www.sapientia.uji.es | 72

UNIVERSITAT JAUME•I

Problemes resolts d'estadística aplicada a les ciències socials

José Modesto J. Beltrán
Pablo Juan Verdoy
María José Peris

Problemes resolts d'estadística aplicada a les ciències socials

Modesto J. Beltrán
Pablo Juan Verdoy
María José Peris

UNIVERSITAT JAUME • I

DEPARTAMENT DE MATEMÀTIQUES

■ Codi d'assignatura RA10, RL0906

Edita: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions
Campus del Riu Sec. Edifici Rectorat i Serveis Centrals. 12071 Castelló de la Plana
http://www.tenda.uji.es e-mail: publicacions@uji.es

Col·lecció Sapientia, 72
www.sapientia.uji.es
Primera edició, 2013

ISBN: 978-84-695-7233-7

Publicacions de la Universitat Jaume I és una editorial membre de l'UNE, cosa que en garanteix la difusió de les obres en els àmbits nacional i internacional. www.une.es

ÍNDEX

Pròleg

L'estadística és una ciència amb base matemàtica referent a la recollida, anàlisi i interpretació de dades, que cerca explicar condicions regulars en fenòmens de tipus aleatori. És transversal a una àmplia varietat de disciplines, des de la física fins a les ciències socials, des de les ciències de la salut fins al control de qualitat, i és usada per a la presa de decisions en àrees de negocis i institucions governamentals. Podem considerar dues branques en l'estadística:

a) L'estadística descriptiva, que es dedica als mètodes de recollida, descripció, visualització i resum de dades originades a partir dels fenòmens en estudi. Les dades poden ser resumides numèricament o gràficament. Exemples bàsics de paràmetres estadístics són: la mitjana i la desviació estàndard. Alguns exemples gràfics són: histograma, piràmide poblacional, agrupament, etc.

b) La inferència estadística, que es dedica a la generació dels models, inferències i prediccions associades als fenòmens en qüestió tenint en compte l'aleatorietat de les observacions. S'usa per a modelar patrons en les dades i extraure inferències sobre la població d'estudi. Aquestes inferències poden prendre la forma de respostes a preguntes si / no (prova d'hipòtesi), estimacions de característiques numèriques (estimació), pronòstics de futures observacions, descripcions d'associació (correlació) o modelització de relacions entre variables (anàlisi de regressió). Altres tècniques de modelització inclouen ANOVA, sèries de temps i mineria de dades.

Ambdues branques (descriptiva i inferencial) comprenen l'estadística aplicada. Hi ha també una disciplina anomenada estadística matemàtica, la qual fa referència a les bases teòriques de la matèria. La paraula «estadística» també es refereix al resultat d'aplicar un algoritme estadístic a un conjunt de dades, com ara estadístiques econòmiques, estadístiques criminals, etc.

En l'origen, per tant, l'estadística va estar associada a dades utilitzades pel govern i cossos administratius (sovint centralitzats). La col·lecció de dades sobre estats i localitats continua àmpliament a través dels serveis d'estadística nacionals i internacionals. En particular, els censos subministren informació regular sobre la població.

Els mètodes estadístico-matemàtics van emergir des de la teoria de probabilitats, la qual data, certament, des de la correspondència entre Pierre de Fermat i Blaise Pascal (1654). Christiaan Huygens (1657) dóna el primer tractament científic que es coneix de la matèria. El *Ars Conjectandi* (pòstum, 1713) de Jakob Bernoulli i la *Doctrina de Possibilitats* (1718) d'Abraham de Moivre estudien la matèria com una branca de les matemàtiques. En l'era moderna, el treball de Kolmogorov ha estat un pilar en la formulació del model fonamental de la teoria de probabilitats, el qual és usat a través de l'estadística. La teoria d'errors es pot remuntar a l'*Òpera Miscellanea* (pòstuma, 1722) de Roger Cotes i al treball preparat per Thomas Simpson en 1755 (imprès en 1756) el qual aplica per primera vegada la teoria de la discussió d'errors d'observació. La reimpressió (1757) d'aquesta obra inclou l'axioma que diu que els errors positius i negatius són igualment probables i que hi ha uns certs límits assignables dins dels quals es troben tots els errors, s'hi descriuen errors continus i una corba de probabilitat.

Pierre-Simon Laplace (1774) fa el primer intent de deduir una regla per a la combinació d'observacions des dels principis de la teoria de probabilitats. Laplace va representar la llei de probabilitats d'errors mitjançant una corba i va deduir una fórmula per la mitjana de tres observacions. També, en 1871, obté la fórmula per a la llei de facilitat de l'error (terme introduït per Lagrange, 1744) però amb equacions immanejables. Daniel Bernoulli (1778) introdueix el principi del màxim producte de les probabilitats d'un sistema d'errors concurrents. *El mètode de mínims quadrats*, el qual va ser usat per minimitzar els errors en mesuraments, va ser publicat independentment per Adrien-Marie Legendre (1805), Robert Adrain (1808), i Carl Friedrich Gauss (1809). Gauss havia usat el mètode en la seua famosa predicció de la localització del planeta nan Ceres el 1801. Proves addicionals van ser escrites per Laplace (1810, 1812), Gauss (1823), James Ivory (1825, 1826), Hagen (1837), Friedrich Bessel (1838), W. F. Donkin (1844, 1856), John Herschel (1850) i Morgan Crofton (1870). Altres com Van Ellis (1844), Augustus De Morgan (1864), Glaisher (1872) i Giovanni Schiaparelli (1875).

El segle XIX inclou autors com Laplace, Silvestre Lacroix (1816), Littrow (1833), Richard Dedekind (1860), Helmert (1872), Hermann Laurent (1873), Liagre, Didion i Karl Pearson. Augustus De Morgan i George Boole milloren la presentació de la teoria. Adolphe Quetelet (1796-1874) va ser un altre important fundador de l'estadística i qui va introduir la noció de l'«home mitjana» (*l'homme moyen*) com un mitjà d'entendre els fenòmens socials complexos com ara taxes de criminalitat, taxes de matrimoni o taxes de suïcidis. Durant el segle XX, la creació d'instruments necessaris per a assumptes de salut pública (epidemiologia, estadística, etc.) i propòsits econòmics i socials (taxa d'atur, econometria, etc.) necessità d'avanços substancials en les pràctiques estadístiques.

Avui l'ús de l'estadística s'ha estés més enllà dels seus orígens com un servei a l'Estat o al govern. Persones i organitzacions usen l'estadística per a entendre dades i prendre decisions en ciències naturals i socials, medicina, negocis i altres àrees. L'estadística és entesa generalment no com un subàrea de les matemàtiques sinó com una ciència diferent, però «aliada». Moltes universitats tenen departaments acadèmics de matemàtiques i estadística separadament. L'estadística s'ensenya en departaments tan diversos com psicologia, educació i salut pública.

El punt de partida de l'aplicació estadística és la preparació d'una població en un problema científic, industrial o social. Aquesta pot ser la població d'un país, la de grans cristal·litzats en una roca o la de béns manufacturats per una fàbrica en particular durant un període concret. També podria ser un procés observat en diversos instants i les dades recollides d'aquesta manera constitueixen una sèrie de temps.

Per raons pràctiques, en lloc de compilar dades d'una població sencera, usualment s'estudia un subconjunt seleccionat de la població, anomenat mostra. Dades sobre la mostra són recollides de manera observacional o experimental. Aleshores, les dades són analitzades estadísticament, la qual cosa cerca dos propòsits: descripció i inferència.

El concepte matemàtic fonamental utilitzat per a entendre l'aleatorietat és el de probabilitat. L'estadística matemàtica (també anomenada teoria estadística) és la branca de les matemàtiques aplicades que fa servir la teoria de probabilitats i l'anàlisi matemàtica per a examinar les bases teòriques de l'estadística. L'ús de qualsevol mètode estadístic és vàlid només quan el sistema o població sota consideració satisfà els supòsits matemàtics del mètode. El mal ús de l'estadística pot produir seriosos errors en la descripció i interpretació, afectant les polítiques socials, la pràctica mèdica i la qualitat d'estructures com ara ponts i plantes de reacció nuclear.

Fins i tot quan l'estadística és correctament aplicada, els resultats poden ser difícilment interpretats per un inexpert. Per exemple, el significat estadístic d'una tendència en les dades, que mesura el grau en què la tendència pot ser causada per una variació aleatòria en la mostra, pot no estar d'acord amb el sentit intuïtiu. El conjunt d'habilitats estadístiques bàsiques (i l'escepticisme) que una persona necessita per a manejar informació diàriament hi és referit com a cultura estadística.

Aquest llibre de problemes amb ajudes és la primera part d'un conjunt de dos que comprendrà totes les fases del procés estadístic. En aquest volum s'estudien mitjançant problemes els principals trets de l'estadística descriptiva d'una variable, de dues variables, dels nombres índexs i de les sèries temporals.

La novetat que presenta aquest manual és que tots els exercicis tenen dos tipus d'ajudes que aporten « pistes» de com resoldre els exercicis i els problemes. Així, l'alumne pot consultar-les sempre que no sàpia per on continuar mentre està resolent un exercici. D'aquesta manera l'estudiant evitarà la desagradable sensació que una persona té quan abandona la resolució d'un exercici. A més a més, també s'hi inclouen les solucions completes dels exercicis, molts d'ells comentats amb profunditat.

És convenient deixar clares dues qüestions rellevants. La primera d'elles és que no s'ha de treure la falsa idea d'entendre l'estadística com una mera col·lecció de mètodes o tècniques útils per al tractament de la informació o, fins i tot el que és més, concloure que l'estadística és el que fan els estadístics. Encara que aquestes dues idees no són desencertades, tampoc no permeten tenir una visió completa del que és l'estadística. La segona és que les nostres decisions es basen, cada vegada més, en un flux creixent d'informació que necessitem sintetitzar per a evitar allò de què els arbres impedeixen veure el bosc. Les nostres decisions són de tipus condicionat, ja que s'hi prenen en funció d'algun tipus d'informació, tant passada com present.

Aquest llibre pretén ser un complement didàctic de la teoria bàsica d'estadística que apareix en molts altres llibres que avui en dia es poden trobar a les nostres biblioteques, i també sobre tot al manual *Introducció a l'estadística aplicada a les ciències socials* de la Col·lecció Sapientia, el qual es pot considerar el manual teòric que complementa aquest llibre.

En definitiva, la nostra humil pretensió és que aquest text servisca d'ajuda complementària a tot l'estudiantat que s'enfronte (moltes vegades amb poc èxit) a la resolució de problemes d'estadística descriptiva.

Els autors
Castelló, gener de 2013

Introducció

El present llibre de problemes es pot considerar com el primer dels dos complements del manual *Introducció a l'estadística aplicada a les ciències socials* de la Col·lecció Sapientia que edita Publicacions de la Universitat Jaume I, el qual consta fonamentalment de continguts teòrics, i amb l'apartat de problemes en un segon pla. Amb aquest nou text, basat quasi exclusivament en problemes resolts, es completa part del manual teòric i es facilita a l'estudiantat una eina excel·lent per a consolidar l'aprenentatge dels seus continguts.

Els problemes compten amb ajudes, sent l'última la seua resolució completa. És a dir, cada un dels problemes té dos tipus d'ajudes, que no són més que una breu informació que pot facilitar a l'estudiant l'ardu treball de resoldre el problema. Les ajudes de *tipus 1* són una mera orientació que tenen com a objecte manifestar els continguts que s'han de consultar per a poder resoldre el problema. L'ajuda de *tipus 2* dóna bastant més informació que la primera. Així, en moltes ajudes d'aquest tipus es mostra part de la resolució de l'exercici. Finalment, en la resolució del problema es mostren fil per randa els continguts estadístics que s'utilitzen i nombrosos comentaris que permeten intuir la resolució de problemes similars.

A més a més, els problemes estan classificats per objectius, ja que d'aquesta manera l'estudiant sap en cada moment quins continguts s'estan treballant, i per tant, pot consultar el manual teòric per tal de revisar aquelles qüestions que presenten dificultats.

Per una altra part, aquest manual està dividit en quatre unitats, que fan referència a l'estadítica descriptiva univariant, l'estadística descriptiva bivariant, els nombres índexs i, finalment, les sèries temporals. Cada unitat està dividida en quatre blocs: en el primer es proposen els enunciats dels problemes classificats per objectius. La segona part proporciona únicament les ajudes de tipus 1. En el tercer bloc, les ajudes de tipus 2. El fet que per a un mateix problema no es troben els dos tipus d'ajudes conjuntament té la pretensió que l'estudiant realitze la consulta detallada de les ajudes, reforçant la idea de *pensar abans de consultar*. En la darrera part es mostren les resolucions completes dels problemes, les quals estan repletes de comentaris, gràfics i diagrames que en faciliten la seua comprensió.

UNITAT 1

Estadística descriptiva univariant

Introducció teòrica

Com a elements introductoris d'aquest capítol, és convenient recordar definicions d'elements importants, ja desenvolupants en diferents materials com ara els llibres referenciats 1, 2 i 3:

Població: És el conjunt d'elements, individus o subjectes a estudi i dels quals es vol obtenir un resultat.

Paràmetre: És una mesura descriptiva de la població total, de totes les observacions.

Mostra: Conjunt d'elements que formen part de la població total a la qual representa.

Grandària de la mostra: És el nombre d'elements o observacions que formen la mostra.

Estadístic: És una mesura descriptiva de la mostra i que estima el paràmetre de la població.

Variables qualitatives i quantitatives

Les variables en què únicament és possible un recompte del nombre d'elements de la població o mostra que posseeixen una de les seues modalitats s'anomenen variables qualitatives o atributs (llibres referenciats 4, 8, 14 i 19). Les modalitats d'aquests tipus de variables ni tan sols admeten una gradació i molt menys una mesura numèrica. Són variables com el sexe d'una persona, la confessionalitat, etc. Les modalitats que poden prendre s'anomenen categories. Així, les categories de la variable sexe són masculí i femení.

La resta de variables en què, a més d'admetre el recompte del nombre d'elements de la població o mostra que posseeixen una de les seues modalitats, també és possible assignar-li una mesura a la mateixa modalitat, s'anomenen variables quantitatives. Són per exemple el pes, l'alçada, el sou mensual, el grau de duresa, etc.

Aquestes darreres variables, les quantitatives, també poden classificar-se en discretes i contínues. Una variable contínua és aquella que pot prendre qualsevol valor dintre d'un rang donat. Independentment de la proximitat de dues observacions, si l'instrument de mesura és suficientment precís, sempre s'hi podrà trobar una tercera observació entre les dues primeres.

Una variable discreta està limitada per a certs valors, generalment nombres enters. Es diferencien de les quantitatives en què, donades dues observacions suficientment pròximes, no es pot trobar cap observació de la variable entre elles. En són exemples el nombre de fills de les famílies, el nombre de vehicles que tenen les empreses, nombre de turistes que visiten un país, etc.

La variable estadística es denota per les majúscules. Així mateix, cada una d'aquestes variables pot prendre distints valors sent la seua notació la següent:

$$X = (x_1, x_2, x_3, \ldots, x_{k-2}, x_{k-1}, x_k)$$

Taules de freqüència

Abans de construir les taules de freqüències, cal realitzar una sèrie de definicions:

S'anomena *freqüència absoluta del valor* x_i al nombre de vegades que apareix repetida l'observació en el recull de dades. Es representa per n_i.

S'anomena *freqüència relativa del valor* x_i al quocient entre la freqüència absoluta d' x_i i el nombre total de dades n. Es representa per f_i i evidentment, és la proporció en què es troba el valor x_i dintre del conjunt de dades en tant per u; $f_i = \frac{n_i}{n}$.

D'altra banda, suposant que es disposarà de k dades diferents, es compleix que la suma de tots els n_i és n ($n_1 + n_2 + \ldots\ldots + n_k = n$), i també que la suma de les freqüències relatives és igual a la unitat ($f_1 + f_2 + \ldots\ldots + f_k = 1$).

S'anomena *freqüència absoluta acumulada del valor* x_i al nombre de dades del recull que són menors o iguals que x_i. Es representa per N_i i el seu valor es calcula a partir de les freqüències absolutes; $N_i = n_1 + n_2 + \ldots\ldots + n_i$ (assumint que $x_1 < x_2 < \ldots\ldots < x_i$).

S'anomena *freqüència relativa acumulada del valor* x_i al quocient entre la freqüència absoluta acumulada d' x_i i el nombre total de dades n. Es representa per F_i i evidentment, és la proporció en què es troben el valors menors o iguals a x_i dintre del conjunt de dades en tant per u; $F_i = \frac{N_i}{n}$. També hi ha una altra manera de calcular F_i a partir de les freqüències relatives, per tant $F_i = f_1 + f_2 + \ldots\ldots + f_i$. (assumint que $x_1 < x_2 < \ldots\ldots < x_i$).

Les freqüències acumulades també compleixen dues propietats trivials com a conseqüència de les seues definicions: suposant que es disposara de k dades diferents, es compleix que $N_k = n$ i $F_k = 1$.

És important remarcar que per a calcular freqüències acumulades és necessari que les variables a estudiar siguen *ordenables*, és a dir, ha de ser possible establir una relació d'ordre entre les variables. En cas contrari, no té cap sentit realitzar els esmentats càlculs.

Aquestes definicions permeten resumir les dades. No obstant, la manera més adequada per a sintetitzar les dades és mitjançant el que es denomina taula de freqüències. En ella apareixen distribuïdes les dades segons les freqüències. Al mateix temps reflecteix tots els conceptes esmentats amb anterioritat.

En ocasions el nombre de dades diferents que s'està estudiant és molt nombrós. Llavors, si es decidira construir una taula com l'anterior, la columna relativa a les x_i seria molt extensa, únicament cal pensar en dues-centes dades diferents dintre d'un recull de quatre-centes.

La solució a aquesta qüestió consisteix a l'agrupament de les dades en intervals o classes, de manera que cada dada pertanga a un sol un interval. En conseqüència, els conceptes relatius a la freqüència que fins ara es referien als valors diferents de les dades, al realitzar l'agrupació han de fer referència als intervals.

Aquesta pràctica, malgrat que ajuda a resumir i aclarir la informació, té en canvi un inconvenient: es perd informació sobre la pròpia distribució de dades. Al agrupar-les en els intervals els valors reals es «difuminen».

Un interval se sol representar per $[L_{i-1}, L_i)$ i es defineix com el conjunt format per tots els valors reals que són majors o iguals que L_{i-1} (*extrem inferior*) i menors que L_i (*extrem superior*).

S'anomena *marca de classe* a la mitjana aritmètica dels dos extrems de l'interval. És evidentment el valor central de l'interval ja que equidista dels extrems. Es denota per c_i. Es calcula $c_i = \dfrac{L_{i-1} + L_i}{2}$.

S'anomena *amplària d'un interval* a la distància que hi ha entre els extrems. Es denota per a_i i es calcula $a_i = L_i - L_{i-1}$.

S'anomena *densitat de freqüència absoluta* d'un interval al quocient entre la freqüència absoluta de l'interval i la seua amplària. Es denota per *di.* Es calcula $d_i = \dfrac{n_i}{a_i}$.

Malgrat tot, en la literatura matemàtica és possible trobar diverses regles per a calcular el nombre adequat d'intervals a partir del nombre de dades, com que no pot superar el 10% del nombre total de dades, com el mètode de l'arrel. Segons aquest mètode, el nombre de classes és igual a l'arrel quadrada del nombre de dades:

$$\text{Nombre classes} = \sqrt{nombre\ \ de\ \ dades}$$

S'anomena *recorregut d'un conjunt de dades* a la diferència entre el valor més gran i el més petit del conjunt. Es denota *Re*.

En conseqüència, l'amplària = $\dfrac{recorregut}{nombre\ de\ classes}$

Coneixent el nombre d'intervals i l'amplària es poden construir fàcilment tots els intervals. En finalitzar la construcció de tots els intervals és necessari comprovar que *totes* les dades pertanyen a un sol interval. Si no es així, cal realitzar alguna modificació en l'amplària o en el nombre d'intervals.

Gràfics estadístics

Els gràfics també són molt útils per a descriure els conjunts de dades (referències 15, 20 i 23). De fet, un gràfic estadístic permet formar-se una primera idea de la distribució de les dades tan sols en una observació. No obstant, cal anar amb compte perquè en algunes ocasions els gràfics presenten «tendències» no atribuïbles al quefer matemàtic.

Diagrama de sectors o diagrama circular: És un cercle dividit en diferents sectors. L'àrea de cada sector és proporcional a la freqüència que es vulga representar, siga absoluta o relativa.

Per a calcular l'angle associat a cada freqüència s'aplica una simple proporció: l'angle associat a una freqüència absoluta n_i és igual a $f_i \cdot 360º$ ($f_i = \frac{n_i}{n}$). Per a la freqüència absoluta acumulada es raonaria de la mateixa manera.

Diagrama de barres: S'utilitza per a representar les dades que no estan agrupades. Consisteix a col·locar sobre un eix horitzontal els diferents valors que pren la variable estadística, i sobre cadascun d'ells alçar un rectangle d'alçària igual a la freqüència (del tipus que s'estiga representant). Tots els rectangles han de tenir la mateixa amplària.

Histogrames: S'utilitzen per a representar dades agrupades en intervals. Consisteix a col·locar sobre un eix horitzontal els diferents intervals. Sobre cadascun d'ells es construeix un rectangle de *superfície* igual a la freqüència que s'estiga representant. Així, les altures dels rectangles han de ser les densitats dels intervals. Cal notar, que a l'eix horitzontal apareixen reflectides les marques de les classes.

Polígon de freqüències: És menys utilitzat que els diagrames de barres i els histogrames, però poden substituir-los. Consisteix a unir mitjançant línies poligonals els extrems superiors de les barres si es tracta de dades sense agrupar, o el punt mitjà de la base superior dels rectangles, si es tracta d'histogrames.

Pictograma: Se sol utilitzar per a expressar un atribut. Se solen utilitzar icones que s'identifiquen amb la variable (exemple una bombeta) i la seua grandària és proporcional a la freqüència.

Mesures de posició

Són coeficients que tracten de representar una determinada distribució, poden ser de dos tipus, centrals i no centrals.

Centrals

Mitjana aritmètica

És el valor que habitualment es pren com a representació de les dades. És la suma de tots els valors de la variable dividida entre el nombre total d'elements. Si les dades estan agrupades, es pren la marca de la classe com a representant de l'interval i es realitzen tots els càlculs com si els valors de la variable foren les marques de les classes.

Si es considera una variable estadística X que té k valors diferents, els quals es representen per x_i i les seues freqüències per n_i, llavors la mitjana aritmètica es calcula:

$$\text{Mitjana aritmètica: } \frac{n_1x_1 + n_2x_2 + \ldots\ldots + n_kx_k}{n}$$

La mitjana aritmètica compleix les següents propietats:

- La suma de les desviacions dels valors de la variable respecte a la mitjana aritmètica és 0.

- Si a tots els valors de la variable se li suma una mateixa constant, la mitjana aritmètica queda augmentada amb aquesta constant.

- Si tots els valors de la variable es multipliquen per una mateixa constant la mitjana aritmètica queda multiplicada per l'esmentada constant.

- Si una variable Y és transformació lineal d'una altra variable X ($Y = a \cdot X + b$; a i b nombres reals), la mitjana aritmètica de Y segueix la mateixa transformació lineal respecte a la mitjana aritmètica de X. És a dir: $\overline{Y} = a \cdot \overline{X} + b$.

- Si en un conjunt de valors es poden obtenir 2 o més subconjunts disjunts que suposen una partició del conjunt total de valors, la mitjana aritmètica del conjunt es relaciona amb la mitjana aritmètica de cada un dels subconjunts disjunts de la forma següent: $\overline{X} = \frac{\sum \overline{X}_i \cdot N_i}{n}$ (sent $\overline{X}_i$ la mitjana de cada subconjunt i N_i el nombre d'elements de cada subconjunt).

Mitjana aritmètica ponderada

De vegades, no tots els valors de la variable tenen el mateix pes. És a dir, cadascun dels valors que pren la variable té assignat un nombre que indica la seua importància, el qual és independent de la pròpia freqüència absoluta.

El càlcul de la mitjana aritmètica ponderada en aquestes casos segueix la següent expressió, on w_i és el pes associat a cada valor de la variable x_i.

$$\overline{X}_W = \frac{\sum_{i=1}^{k} x_i w_i n_i}{\sum_{i=1}^{k} w_i n_i}$$

Mitjana geomètrica

Pot utilitzar-se per a mostrar canvis percentuals en una sèrie de nombres positius. Per tant, té una amplia aplicació en els negocis i en l'economia. La mitjana geomètrica proporciona una mesura precisa d'un canvi percentual mitjà en una sèrie de nombres. Es representa per *G* i el seu càlcul –seguint la notació habitual– segueix l'expressió següent:

$$G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot ...... x_k^{n_k} \cdot}$$

Emprant la notació potencial, també es pot representar com:

$$G = (x_1^{n_1} \cdot x_2^{n_2} \cdot ...... x_k^{n_k})^{\frac{1}{n}}$$

Mitjana harmònica

Es representa per *H* i és la inversa de la mitjana aritmètica de les inverses dels valors de la variable, respon a l'expressió següent:

$$H = \frac{n}{\sum \frac{n_i}{x_i}} = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + ... + \frac{n_k}{x_k}}$$

$n =$ *nombre total de dades*

$x_i =$ *valors diferents que pren la variable*

$n_i =$ *freqüència absoluta de* x_i

S'utilitza per a calcular el valor mitjà de magnituds expressades en termes relatius com velocitats, temps, rendiment, tipus de canvi monetari, etc. La seua principal contrarietat és que quan algun valor de la variable és 0 o pròxim a zero no es pot calcular.

En moltes ocasions, no és necessari aplicar la fórmula anterior. Únicament cal tenir present el concepte de mitjana aritmètica.

Mediana

La mediana és el valor de la variable que divideix les observacions en dos grups d'igual nombre d'elements, de manera que en el primer grup totes les dades siguen menors o iguals que la mediana, i en l'altre grup, totes les dades hi siguen majors o iguals. Per tant, és una quantitat que indica *ordre* dins de l'ordenació.

Dades no agrupades

En ordenar les dades, la posició que ocupa la mediana es determina dividint el nombre total de valors entre 2 ($\frac{n}{2}$) o allò que és el mateix, calculant el 50% del total de dades (0,5 · n). Cal tenir en compte però, la paritat de *n*:

- Quan hi haja un nombre senar de valors , la mediana serà just el valor central. Si hi ha moltes dades el càlcul no és immediat, cal construir la taula de freqüències i fixar-se en la columna de les freqüències absolutes acumulades N_i. La mediana serà el valor de variable que tinga la freqüència absoluta acumulada igual a $\frac{n}{2}$. És a dir:

$$si\ N_{i-1} \leq \frac{n}{2} \leq N_i \rightarrow Me = x_i$$

- Quan hi haja un nombre parell de valors, la mediana serà la mitjana aritmètica dels dos valors centrals de la variable. De la mateixa manera que en el cas anterior, si el conjunt d'observacions és nombrós, és necessari construir la taula de freqüències i fixar-se en la columna de les N_i. Si al calcular $\frac{n}{2}$ aquest resulta ser un valor menor que una freqüència absoluta acumulada, la mediana es calcularà de la mateixa manera que en el cas anterior; és a dir $si\ N_{i-1} \leq \frac{n}{2} \leq N_i \rightarrow Me = x_i$. Tanmateix, si coincideix $\frac{n}{2}$ amb algun N_i, per a obtenir-la es realitzarà el càlcul següent: $Me = \frac{x_i + x_{i+1}}{2}$. Els exemples següents aclareixen els càlculs.

Dades agrupades

En distribucions agrupades, cal determinar l'interval [Li-1, Li) en el qual es troba la mediana. Aquest interval es determina seguint exactament els mateixos procediments esmentats en l'apartat anterior; es realitza el mateix que en el cas de dades no agrupades. La diferència radica en què s'obtindrà un interval en lloc d'un valor. Una vegada es té l'interval [L_{i-1}, L_i) , la mediana es calcula:

$$Me = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i \ \text{ on,}$$

L_{i-1} Límit inferior de la classe mediana
N_{i-1} És la freqüència absoluta acumulada de la classe «anterior» a la classe mediana
n_i És la freqüència de la classe mediana
a_i És l'amplària de la classe mediana

És evident que es pretén calcular un representat de l'interval amb l'objecte de fixar la mediana en un valor. Una possibilitat haguera estat considerar la marca de classe, no obstant, el criteri usualment més seguit no és aquest sinó el de la fórmula abans esmentada.

En aquesta fórmula es considera, en primer lloc, el supòsit que les dades estan uniformement distribuïdes dintre de cada interval. Tenint en compte aquest fet, es pot observar que la fórmula és una relació de proporcionalitat entre les posicions que ocupen els valors de la variable i l'amplària dels intervals.

Moda

És el valor de la variable que més vegades es repetix, és a dir, el valor que té major freqüència absoluta.

Poden existir distribucions amb més d'una moda: bimodals, trimodals, etc.

Dades no agrupades

En les distribucions sense agrupar, l'obtenció de la moda és immediata.

Dades agrupades

En els supòsits que la distribució vinga donada en intervals, es poden produir dos casos: que hi tinguen la mateixa amplària, o que aquesta siga distinta. En ambdós casos l'objectiu es trobar un valor que represente la moda.

Intervals amb la mateixa amplària

És evident que una vegada determinada la major freqüència a aquesta no li correspon un valor sinó un interval. Llavors no tindrem un valor modal sinó un interval modal. Per a calcular el representat de l'interval que faça el paper de moda hi ha distints criteris. En el text es recull el següent. En primer lloc es calcula l'interval on es troba la moda, és a dir, l'interval modal $[L_{i-1}, L_i)$ el qual té major freqüència absoluta (n_i). Posteriorment es calcula la moda de la següent manera:

$$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} \cdot a_i$$

On:

L_{i-1}: extrem inferior de l'interval modal
a_i: amplitud de tal interval
n_{i-1} , n_{i+1}: freqüències dels intervals anterior i posterior respectivament de l'interval modal

De la mateixa manera que la mediana, la fórmula té el supòsit que les dades estan uniformement repartides dintre de cada interval. A més a més, seguint aquest criteri es pot observar que la moda estarà més a prop d'aquell interval adjacent amb major freqüència absoluta.

No Centrals

Són mesures de localització semblants a la mediana. La seua funció és informar del valor de la variable que ocuparà la posició (en tant per cent) que ens interesse respecte de tot el conjunt d'observacions.

Podem dir que els quantils són unes mesures de posició que dividixen a la distribució en un cert nombre de parts.

Les més importants són:

- *Quartils,* divideixen la distribució en quatre parts iguals (tres divisions). C_1, C_2, C_3, corresponents al 25%, 50%, 75%. Per exemple, el 1r quartil té un 25 % de les dades menors o iguals a ell, el segon quartil és la mediana, etc.

- *Decils,* divideixen la distribució en 10 parts iguals (9 divisions). D_1,..., D_9, corresponents al 10%,..., 90%.

- *Percentils,* divideixen la distribució en 100 parts (99 divisions). P_1,...,P_{99}, corresponents a l'1%,..., 99%. Per exemple, el valor corresponent al percentil 65, té un 65% de les dades menors o iguals a ell.

Hi ha un valor en qual coincidixen els quartils, els decils i percentils. És la mediana, ja que: $P_{50} = C_2 = D_5$. El càlcul del quantils segueix el mateix procediment que el que s'ha utilitzat en la mediana, tant per a les dades agrupades com per a les dades sense agrupar. Així, en general es calcula la posició en què es troba el quantil i després s'hi calcula. Es distingeix entre distribucions agrupades, i les que no ho estan:

Dades no agrupades

En primer lloc es calcula la posició que ocupa el quantil que s'està calculant. Així, determina si Q_a representa el quantil que deixa per sota d'ell un a (%) de les dades:

$$si\ N_{i-1} \leq \frac{a}{100} \cdot n \leq N_i \rightarrow Q_a = x_i$$

$$\text{en el supòsit que } \frac{a}{100} \cdot n = N_i \rightarrow Q = \frac{x_i + x_{i+1}}{2}$$

DADES AGRUPADES

En distribucions agrupades, cal determinar l'interval $[L_{i-1}, L_i)$ en el qual es troba el quantil. Aquest interval es determina seguint exactament els mateixos procediments esmentats a l'apartat anterior; es realitza el mateix que en el cas de dades no agrupades. La diferència radica en què s'obtindrà un interval en lloc d'un valor. Una vegada es té l'interval $[L_{i-1}, L_i)$, el quantil es calcula:

$$Me = L_{i-1} + \frac{\frac{a}{100} \cdot n - N_{i-1}}{n_i} a_i \text{ on,}$$

L_{i-1} Límit inferior de la classe mediana
N_{i-1} És la freqüència absoluta acumulada de la classe «anterior» a la calsse mediana
n_i És la freqüència de la classe mediana
a_i És l'amplària de la classe mediana

Mesures de dispersió

Són complementàries de les de posició, en el sentit que assenyalen la dispersió del conjunt de totes les dades de la distribució, respecte de la mesura o mesures de localització adoptades.

Recorregut

Es definix com la diferència entre el major i menor valor de les variables d'una distribució de dades, és a dir:

$$\mathrm{Re} = \max(x_i) - \min(x_i)$$

Recorregut intercuartíl.lic

Es definix com la distància que hi ha entre el tercer i el primer quartil, és a dir:

$$\mathrm{Re} = C_3 - C_1$$

Desviació mitjana respecte de la mediana

És definix com la mitjana aritmètica dels valors absoluts de les desviacions dels valors de la variable respecte de la mediana. Respon a l'expressió següent:

$$D_{|Me|} = \frac{\sum_{i=1}^{k} |x_i - Me| \cdot n_i}{n}$$

Variància

Es defineix com la mitjana aritmètica dels quadrats de les desviacions dels valors de la variable respecte de la mitjana aritmètica de la distribució. Respon a l'expressió:

$$s^2 = \frac{(x_1 - \overline{X})^2 \cdot n_1 + (x_2 - \overline{X})^2 \cdot n_2 + \ldots + (x_k - \overline{X})^2 \cdot n_k}{n} = \frac{\sum_{i=1}^{k} (x_i - \overline{X})^2 \cdot n_i}{n}$$

Com es pot observar, la variància es una mitjana del quadrat dels errors que es cometen en considerar la mitjana aritmètica com «el representant» de totes i cada una de les dades.

D'altra banda, una de les principals dificultats que presenta la variància és la unitat, ja que ve donada en unitats al quadrat (h^2, m^2, etc.). La solució a aquesta circumstància és calcular l'arrel quadrada.

Desviació típica o desviació estàndard

Es defineix com l'arrel quadrada, amb signe positiu, de la variància. Respon a la següent expressió:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{X})^2 \cdot n_i}{n}}$$

En les definicions anteriors s'hi han considerat dades no agrupades. Si ho foren, únicament caldria emprar les marques de classes com a representats dels intervals. És a dir, $c_i = x_i$.

Per una altra part, es poden definir dos estadístics de dispersió més, anomenats quasivariància i quasidesviació típica com:

$$s^2_{n-1} = \frac{n}{n-1} s^2 \quad i \quad s_{n-1} = \sqrt{\frac{n}{n-1}} \cdot s$$

Aquests estadístics tenen molt d'interés en l'estadística inferencial com es veurà en capítols posteriors.

La variància compleix les propietats següents:

- La variància és sempre un valor no negatiu o 0. Únicament pot ser 0 si totes les dades són iguals. En aquest cas és evident que $\overline{X} = x_i$ per a tot els possibles valors de l'índex.

- Si a tots els valors de la variable se'ls suma una constant, la variància no es modifica.

- Si tots els valors de la variable es multipliquen per una constant, la variància queda multiplicada pel quadrat de dita constant.

- Si una variable *X'* és transformació lineal d'una altra variable *X* ($X' = a \cdot X + b$; a i b nombres reals), la variància de *X'* s'obté a partir de la de *X* del mode $s'^2 = a^2 \cdot s^2$.

- Si d'un conjunt de valors es poden obtenir dos o més subconjunts disjunts que formen una partició d'aquest conjunt, la variància de tot el conjunt de dades està relacionada amb les variàncies dels subconjunts.

- Les mesures de dispersió absolutes són uns indicadors que presenten dificultats a l'hora de comparar la representativitat de les mesures de tendència central entre dues distrubucions de dades diferents. Per això, a vegades es recorre a mesures de dispersió relatives. El coeficient de variació de *Pearson* és una de les més significatives i determina el grau de significació d'un conjunt de dades relatiu a la seua mitjana aritmètica. Es defineix com el quocient entre la desviació típica i la mitjana aritmètica de la distribució de dades.

$$V_X = \frac{s}{\left|\overline{X}\right|}$$

Mesures de forma

Ens donen informació de la forma de l'histograma, de la seua simetria i de la menor o menor proximitat dels valors de la variable respecte de la seua mitjana.

Coeficient d'asimetria de Fisher

Les mesures d'asimetria permeten determinar, sense que siga necessari fer les representacions gràfiques, el grau de simetria que presenten les dades respecte d'un valor central de la variable estadística, normalment la mitjana aritmètica. Per tant, aquesta mesura ha de reflectir dos aspectes: la distància de cada observació respecte a la mitjana aritmètica (és a dir, la diferència entre cada valor i la mitjana aritmètica: $x_i - \overline{x}$) i la freqüència de cadascuna d'aquestes distàncies (la qual coincidirà, evidentment, amb la freqüència de cada observació). D'aquesta manera, intuïtivament, si «predominen» les distàncies negatives sobre les positives (per ser més freqüents o ser distàncies molt grans), la distribució és asimètrica a esquerres. Si pel contrari, es dóna la situació oposada, llavors la distribució és asimètrica a dretes. Per acabar, si les distàncies negatives i les positives se «compensen», la distribució és simètrica.

Ara doncs, el que cal és trobar l'estadístic que determine la asimetria de la distribució de dades. Com que hi està directament relacionada amb les desviacions respecte a la mitjana aritmètica, una primer apropament pot ser la mitjana de les desviacions, és a dir, $\frac{\sum_{i=1}^{k}(x_i - \overline{X})n_i}{n}$. No obstant, ja és conegut que aquesta suma és zero (propietats de la mitjana aritmètica).

D'altra banda, com que ens interessa conèixer el signe de les desviacions, tampoc no podem emprar el quadrat de les desviacions. Així doncs, sembla coherent prendre una potència de grau tres de les desviacions i calcular-ne la mitjana. Així, si anomenem:

$$\mathrm{m} = \frac{\sum_{i=1}^{k}(x_i - \overline{X})^3 n_i}{n},$$

llavors es compleix:

si $m = 0$	*la distribució és simètrica*
si $m > 0$	*la distribució és asimètrica positiva*
si $m < 0$	*la distribució és asimètrica negativa*

D'aquesta manera s'obté el *coeficient d'asimetria de Fisher*.

$$g_1 = \frac{m}{s^3} = \frac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^3 n_i}{n}}{\left(\sqrt{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^2 n_i}{n}}\right)^3}$$

Curtosi

Per a estudiar el grau de curtosi d'una distribució cal prendre un model teòric com a referència, la representació gràfica del qual tinga forma de campana simètrica. No és estrany doncs, que es prenga el model normal, ja que, com ja s'ha esmentat amb anterioritat, es pot dir que és el model campaniforme per antonomàsia.

D'aquesta manera, prenent aquest model com a referència es diu que una distribució és *leptocúrtica* si és més apuntada que la distribució normal. Si ho és menys, se l'anomena *platicúrtica*. Finalment, si té el mateix apuntament que una distribució normal se la denomina *mesocúrtica*.

De la mateixa manera que en el cas de l'estudi de la asimetria, hi ha un coeficient que permet classificar les dades segons la curtosi. En aquest cas, el coeficient no és tan intuïtiu, i per això únicament es donarà la definició i la seua interpetació. Com en el cas de l'altra mesura de forma, aquest indicador tampoc té dimensió.

$$g_2 = \frac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^4 n_i}{n}}{\left(\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^2 n_i}{n}\right)^2} - 3$$

La idea de l'apuntament d'una distribució de dades surt de la comparació de la freqüència dels valors centrals d'una distribució amb la freqüència dels valors centrals en un model teòric normal que tinga la mateixa mitjana i la mateixa desviació típica que la distribució que s'estudia.

Com que en un model normal es compleix que $\dfrac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^4 n_i}{n}}{s^4} = 3$, llavors:

Una distribució serà:

mesocúrtica (normal)	si $g_2 = 0$
leptocúrtica	si $g_2 > 0$
platicúrtica	si $g_2 < 0$

Per últim, cal remarcar que l'estudi de la curtosi no implica necessàriament que les distribucions siguen simètriques. Així, per exemple, ens podríem trobar distribucions d'observacions que siguen leptocúrtiques i, al mateix temps, asimètriques positives.

Caixes i bigot (Box-Plot)

Un diagrama de caixes i bigot (conegut també com *Box and whisker Plot* en anglès), és una representació gràfica de les dades que permet determinar amb molta facilitat i d'una manera visual la tendència central, la variabilitat, la asimetria i l'existència de valors anòmals d'un conjunt d'observacions. D'alguna manera, es pot dir que és un dels gràfics que més i millor resumeixen els conjunts de dades.

El diagrama de caixes empra el resum dels 5 nombres: la menor observació, la major observació, el primer quartil, la mediana i el tercer quartil.

Mesures de concentració

Estudien el grau de concentració d'una magnitud, normalment econòmica, en determinats individus. En certa manera és un terme oposat a l'equitat en el repartiment. Es denomina *concentració* al grau d'equitat en el repartiment de la suma total dels valors de la variable considerada (renda, salaris, etc.).

Les infinites possibilitats que poden adoptar els valors, es troben entre els dos extrems:

Concentració màxima, quan un només percep el total i els altres res, en aquest cas, s'està davant d'un repartiment no equitatiu:

$$el\ que\ rep\ x_1 = el\ que\ rep\ x_2 = \ldots\ldots = el\ que\ rep\ x_{k-1} = 0\ \ i\ \ el\ que\ rep\ x_k = el\ total$$

Concentració mínima, quan el conjunt total de valors de la variable esta repartit per igual, en aquest cas s'està davant d'un repartiment equitatiu:

$$el\ que\ rep\ x_1 = el\ que\ rep\ x_2 = \ldots\ldots = el\ que\ rep\ x_{k-1} = el\ que\ rep\ x_k$$

Hi ha diferents mesures de concentració, però en el text s'estudiarà l'índex de Gini; per ser un coeficient, serà un valor numèric. Per a obtenir-lo és necessari realitzar un conjunt de càlculs.

Se suposa que hi ha una distribució de rendes ($x_i \cdot n_i$) on i pren els valors d'1 fins a $\boldsymbol{k}$ (per exemple, x_i *són els sous* i n_i el nombre de persones que cobren aquest sou) de la qual es formarà una taula amb les columnes següents:

1) Els productes, $x_i \cdot n_i$ indicaran la renda total percebuda pels n_i rendistes de renda individual x_i.
2) Les freqüències absolutes acumulades N_i.
3) Els totals acumulats u_i que es calculen de la forma següent:

$$\begin{aligned}
\boldsymbol{u}_1 &= x_1 n_1 \\
\boldsymbol{u}_2 &= x_1 n_1 + x_2 n_2 \\
\boldsymbol{u}_3 &= x_1 n_1 + x_2 n_2 + x_3 n_3 \\
\boldsymbol{u}_4 &= x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\boldsymbol{u}_k &= x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4 + \ldots\ldots\ldots\ldots + x_k n_k
\end{aligned}$$

Per tant, es pot dir que:

$$u_j = \sum_{i=1}^{j} x_i \cdot n_i \text{ per a qualsevol valor de } j \text{ des d'1 fins a } k.$$

4) La columna total de freqüències acumulades relatives, que s'expressa en tant per cent i que es representa per p_i, vindrà donada per la següent notació:

$$p_i = \frac{N_i}{n}$$

5) La columna de renda acumulada relativa, que s'expressa en tant per cent i que es representa per l'expressió:

$$q_i = \frac{u_i}{u_k}$$

Aleshores ja es pot confeccionar la taula:

x_i	n_i	$x_i n_i$	N_i	u_i	$p_i = \frac{N_i}{n}$	$q_i = \frac{u_i}{u_k}$	pi - qi
x_1	n_1	$x_1 n_1$	N_1	u_1	p_1	q_1	$p_1 - q_1$
x_2	n_2	$x_2 n_2$	N_2	u_2	p_2	q_2	$p_2 - q_2$
...	...	...	...	...	...	...	...
x_k	n_k	$x_k n_k$	N_k	u_k	100	100	0

Com es pot veure, l'última columna és la diferència entre les dues penúltimes; aquesta diferència seria 0 per a la concentració mínima en la qual es compleix $p_i = q_i$ per a qualsevol i, per tant la seua diferència seria zero.

Analíticament l'índex de Gini: $I_G = \dfrac{\sum_{j=1}^{k-1} (p_i - q_i)}{\sum_{j=1}^{k-1} p_i}$

Aquest índex prendrà els valors:

- $i_G = 0$ quan $p_i = q_i$ concentració mínima
- $i_G = 1$ quan $q_i = 0$ concentració màxima

Per altra banda, si es representen gràficament els q_i a l'eix vertical i els p_i a l'horitzontal s'obtindrà la **corba de concentració o corba de Lorenz.** Es pot comprovar que aquesta corba resultant sempre apareixerà «per sota» de la diagonal del primer quadrant, la qual representa la concentració mínima. A més a més, com més s'aproxime aquesta corba a la diagonal, menys serà la concentració.

A continuació, es desenvoluparan els objectius i els exercicis corresponents a aquest capítol. Cal recordar que el material desenvolupat i el resultat d'alguns exemples són aplicacions desenvolupades amb el software R (referències bibliogràfiques 13, 18 i 22).

Objectius

Els problemes han de permetre que l'alumnat assolisca els objectius didàctics següents:

1*a*) Conèixer els conceptes bàsics de les variables estadístiques.
1*b*) Saber classificar les variables estadístiques.
1*c*) Saber analitzar i realitzar taules de freqüències d'un conjunt de dades.
1*d*) Conèixer les diferències entre les taules de dades sense agrupar i les taules de dades agrupades.
1*e*) Saber interpretar i construir els principals gràfics estadístics.
1*f*) Conèixer els conceptes i saber realitzar els càlculs de les mesures de tendència central i de dispersió. Concretar amb l'aplicació del coeficient de variació de Pearson en aquelles situacions que ho requerisquen.
1*g*) Conèixer els principals estadístics que mesuren la forma de les dades a partir dels gràfics.
1*h*) Saber calcular i interpretar l'índex de Gini, així com saber realitzar la corba de Lorenz per a mesurar l'equitat d'un repartiment.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

Objectiu Exercici	***1a***	***1b***	***1c***	***1d***	***1e***	***1f***	***1g***	***1h***
1	**x**	**x**						
2	x	x			x			
3	x	x			x			
4		x	x	x				
5		x	x	x				
6		x	x		x	x		
7		x					x	
8								x
9	x				x	x		

Enunciats

1*a*) Conèixer els conceptes bàsics de les variables estadístiques.
1*b*) Saber classificar les variables estadístiques.

Exercici 1

Classifique les variables següents, justificant el perquè de l'elecció:

a) Color dels cotxes.
b) Marques d'ordinadors.
c) Longitud de carreteres en metres.
d) Nivell d'estudis.
e) Nombre de fills d'una família.
f) Nombre d'alumnes d'estadística en una carrera.
g) Metres d'altitud de les muntanyes.
h) Professions de les persones.
i) Sou mensual dels treballadors de les empreses del sector ceràmic.

1*a*) Conèixer els conceptes bàsics de les variables estadístiques.
1*b*) Saber classificar les variables estadístiques.
1*e*) Saber interpretar i construir els principals gràfics estadístics.

Exercici 2

Actualment, s'estudia en les distintes comunitats autònomes el nombre de fills per família per a analitzar la natalitat. Un dels treballadors que està fent les enquestes, arreplega les dades del seu barri on hi ha 100 famílies. Hi ha obtingut les dades següents:

1	3	3	0	4	3	1	4	0	0
2	1	0	3	1	2	1	4	1	2
3	3	4	2	0	4	3	0	2	3
1	3	4	2	2	4	4	4	2	1
4	2	1	1	0	1	1	2	3	0
3	3	3	1	1	3	3	0	2	3
4	3	0	3	1	2	2	1	2	3
3	2	1	3	1	3	4	4	4	1
3	0	3	1	0	4	3	2	3	2
1	2	0	2	0	0	2	2	3	4

a) Construïu el gràfic que consideres més adequat amb les freqüències acumulades.
b) Construïu el polígon de freqüències amb les freqüències acumulades.

1*a*) Conèixer els conceptes bàsics de les variables estadístiques.
1*b*) Saber classificar les variables estadístiques.
1*e*) Saber interpretar i construir els principals gràfics estadístics.

Exercici 3

Els sous, en milers d'euros mensuals de 40 empresaris del sector de la construcció de l'any 2007 són:

3,9	4,7	3,7	5,6	4,3	4,9	5,0	6,1	5,1	4,5
5,3	3,9	4,3	5,0	6,0	4,7	5,1	4,2	4,4	5,8
3,3	4,3	4,1	5,8	4,4	4,8	6,1	4,3	5,3	4,5
4,0	5,4	3,9	4,7	3,3	4,5	4,7	4,2	4,5	4,8

Es vol estudiar si realment són prou alts i quina és la seua distribució. Per aconseguir-ho:

a) Representeu gràficament la informació arreplegada.
b) Creu la mateixa representació en 4 classes per a poder diferenciar de forma més clara els tipus de sous.

1*b*) Saber classificar les variables estadístiques.
1*c*) Saber analitzar i realitzar taules de freqüències d'un conjunt de dades.
1*d*) Conèixer les diferències entre les taules de dades sense agrupar i les taules de dades agrupades.

Exercici 4

El recull de 20 dades corresponents al nombre de telefonades enregistrades en una empresa durant els dies de preparació de material per a una fira de mostres durant el període de 9 a 12 hores:

15 ,5, 10, 5, 5, 6, 5, 6, 5, 6, 7, 10, 10, 12, 11, 11, 12, 15, 12, 15

Es vol estudiar si realment hi ha variació al llarg dels dies de les telefonades que es reben. Per aquest motiu es demana confeccionar una taula de freqüències que reculla aquesta informació.

1*b*) Saber classificar les variables estadístiques.
1*c*) Saber analitzar i realitzar taules de freqüències d'un conjunt de dades.
1*d*) Conèixer les diferències entre les taules de dades sense agrupar i les taules de dades agrupades.

Exercici 5

Una empresa fa l'estudi dels diners que es gasta la gent per comprar una segon casa com a complement de la primera vivenda. Recull les dades dels euros i el nombre de famílies que han comprat aquest tipus de vivenda. A continuació es poden veure les dades:

Euros	Famílies
0-50000	2145
50000-75000	1520
75000-100000	840
100000-115000	955
115000-135000	1110
135000-140000	2342
140000-150000	610
150000-200000	328
>200000	150

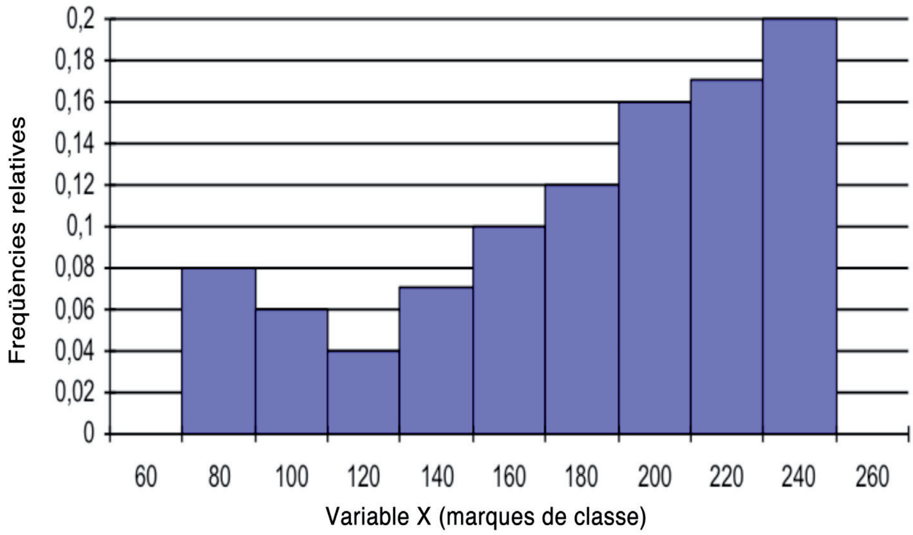
Es demana:

a) De quin tipus de variable és l'objecte d'estudi?
b) Mostreu en forma de taula de freqüències el conjunt de les dades arreplegades.
c) Quin percentatge de famílies es gasten més de 150000 euros?
d) El 65% de famílies que menys es gasta, quina quantitat de diners desemborsa com a màxim?

1*b*) Saber classificar les variables estadístiques.
1*c*) Saber analitzar i realitzar taules de freqüències d'un conjunt de dades.
1*e*) Saber interpretar i construir els principals gràfics estadístics.
1*f*) Conèixer els conceptes i saber realitzar els càlculs de les mesures de tendència central i de dispersió. Concretar amb l'aplicació del coeficient de variació de Pearson en aquelles situacions que ho requerisquen.

Exercici 6

En el següent histograma, es representa la distribució dels diners que durant l'últim mes s'han gastat els treballadors d'una empresa en dietes:



a) Determineu, sabent que hi ha 200 treballadors.
b) La taula de freqüències que mostra les dades que tenim.
c) La quantitat mitjana que s'han gastat, la més freqüent i la quantitat que tenien com a màxim, el 50% dels treballadors que menys cobraven.
d) Calculeu i interpreteu el rang de la distribució així com el rang interquartíl·lic.
e) Calculeu el mínim del 20% dels empleats amb major quantitat de dietes. Quin percentatge del total de l'empresa correspon a aquest grup?
f) L'interval centrat en la quantitat mitjana en què es troben el 75% de les dades. És, doncs, el sou mitjà molt representatiu del conjunt de les dietes?
g) El mes següent, l'empresa decidí augmentar les dietes de tots els treballadors un 5%. A més a més, els hi donà una prima de 50 euros en concepte de productivitat. Calculeu el salari mitjà, el salari més freqüent i el salari que tenien com a màxim, el 50% dels treballadors que menys cobren el mes següent.

h) De les dietes d'una altra empresa, que pertany al mateix sector, se sap que la mitjana aritmètica dels seus treballadors és de 120 euros, amb una variància de 2,5 euros. Quina empresa té una dieta mitjana més representativa? Raoneu la resposta.

1*b*) Saber classificar les variables estadístiques.
1*g*) Conèixer els principals estadístics que mesuren la forma de les dades a partir dels gràfics.

Exercici 7

Es vol llançar al mercat un nou producte ceràmic i l'empresa que el crea estudia el temps de publicitat, en segons, que unes altres empreses han utilitzat per a promocionar un producte similar. A continuació es pot veure, per a cada empresa, la duració i els anuncis realitzats:

Empresa 1

Duració	Nombre d'anuncis
0-20	3
20-25	17
25-30	13
30-40	9
40-60	8

Empresa 2

Duració	Nombre d'anuncis
0-20	1
20-25	5
25-30	13
30-40	5
40-60	2

Empresa 3

Duració	Nombre d'anuncis
0-20	4
20-25	6
25-30	7
30-40	5
40-60	3

Empresa 4

Duració	Nombre d'anuncis
0-20	3
20-25	17
25-30	13
30-40	9
40-60	8

Per a fer l'estudi, calculeu:

a) La durada mitjana de cada empresa.
b) Tenen totes les distribucions la mateixa forma? Comenteu-ne el resultat.

1*h*) Saber calcular i interpretar l'índex de Gini, així com saber realitzar la corba de Lorenz per a mesurar l'equitat d'un repartiment.

Exercici 8

Dues companyies de venda de cotxes tenen maneres diferents de pagar als seus treballadors. La companyia A ho fa mitjançant un sou fix mensual i la companyia B mitjançant un percentatge sobre les vendes efectuades. La distribució dels salaris per categories és la següent:

COMPANYIA A	
Sou (centenes d'euro)	Nombre de treballadors
26	10
39	10
52	40
247	20
260	10
273	10

COMPANYIA B	
Sou (centenes d'euro)	Nombre de treballadors
4	10
5	10
6	40
7	20
26	10
27	10

a) Basant-se únicament en les observacions, en quina companyia el sou mitjà fluctua menys o té els repartiments més equitatius? Justifiqueu-ne el resultat mitjançant l'anàlisi estadístic del repartiment.
b) En quina de les dues companyies el sou és més homogeni o concentrat? S'ha d'obtenir el resultat també de forma gràfica.

1*a*) Conèixer els conceptes bàsics de les variables estadístiques.
1*e*) Saber interpretar i construir els principals gràfics estadístics.
1*f*) Conèixer els conceptes i saber realitzar els càlculs de les mesures de tendència central i de dispersió. Concretar amb l'aplicació del coeficient de variació de Pearson en aquelles situacions que ho requerisquen.

Exercici 9

La distribució d'edats del Cens Electoral de Residents a 1 de gener de 1999 per a les comunitats autònomes d'Aragó i Canàries, en tants per cent és la següent:

Edats	Aragó	Canàries
16–18	3,55	4,35
18–30	21,56	29,99
30–50	31,63	35,21
50–70	28,14	21,97
70–90	15,12	8,48

a) Representeu sobre els mateixos eixos de coordenades les dades de la distribució de l'edat per a les dues comunitats autònomes (empreu distint traç o distints colors). Quines conclusions obteniu a la vista del gràfic?
b) Calculeu l'edat mitjana per a les dues comunitats. Compareu-les. Què indiquen aquests resultats?
c) En quina comunitat les observacions són més disperses?
d) Si les dades d'edats foren: Aragó: 10, 10, 10, 10, 20, 30, 40, 30, 40, 50, 60, 40, 40, 40, 60, 70, 80, 70, 80, 90, 70, 50, 40, 90. Canàries: 20, 30, 40, 40, 140, 50, 40, 30, 40, 30, 50, 60, 40, 30, 30, 40, 30, 40, 30, 40, 30, 50, 60, 70. Obteniu un gràfic que ens mostre la dispersió de les dades.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de realitzar els exercicis i problemes. És convenient no fer un abús excessiu d'aquestes ajudes, és a dir, abans d'emprar-la cal pensar el problema almenys durant uns 10 o 15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; i en darrer terme, la solució.

Ajudes Tipus 1

Exercici 1

El que es necessita per resoldre aquest exercici, és primerament conèixer els tipus de variables que existeixen. A continuació pots veure'n una classificació.

Les *variables qualitatives* són les que no es poden mesurar, és a dir, aquelles que prenen valors als quals no es pot assignar cap número. Expressen qualitats o categories. A més a més, poden ser:

a) *Ordinals*: es poden ordenar.
b) *Nominals*: no hi ha preferències entre unes i altres.

Les *variables quantitatives*, al contrari, són mesurables, és a dir, els valors que s'hi observen poden expressar-se de forma numèrica. Aquestes variables poden classificar-se en:

a) *Discretes*, quan prenen els seus valors en un conjunt finit o numerable.
b) *Contínues*, quan poden prendre qualsevol valor en un interval.

Exercici 2

El que es necessita per resoldre aquest exercici, és conèixer primerament els tipus de variables que existeixen per triar-ne la correcta i el tipus de gràfic corresponent.

La classificació dels tipus de variables, com ja es coneix de l'exercici anterior és:

– Les *variables qualitatives* (*Ordinals o Nominals*).
– Les *variables quantitatives* (*Discretes o Contínues*).

Segons el tipus de variable, el gràfic corresponent serà:

- Per a les *variables qualitatives*: diagrama de barres o diagrama de sectors.
- Per a les *variables quantitatives*:
  Discretes: diagrama de barres o sectors.
  Contínues: histograma.

Els primers passos seran saber quin tipus de variable és, ja que aquest element afectarà a l'elecció tant del tipus de taula de freqüències com a l'elecció del tipus de gràfic.

Queda clar que és una variable numèrica. Per tant, pot ser contínua o discreta. En aquest cas, com que les dades fan referència a nombre de fills serà quantitativa discreta.

Amb aquestes informacions, es pot passar a resoldre el problema.

Exercici 3

El que es necessita per a resoldre aquest exercici, de la mateixa forma que l'anterior, és conèixer els tipus de variables que existeixen per triar-ne la correcta i el tipus de gràfic corresponent, i en aquest cas el que apareixen són dades numèriques continues. Per aquest motiu, el que es treballa és la creació de representacions gràfiques com són els histogrames en els dos apartats.

Per altra part, cal pensar com crear les classes per fer aquest tipus de problemes i es pot fer amb el coneixement dels elements següents:

S'anomena *marca de classe* a la mitjana aritmètica dels dos extrems de l'interval. És evidentment el valor central de l'interval ja que equidista dels extrems. Es denota per c_i. Es calcula $c_i = \frac{L_{i-1} + L_i}{2}$.

S'anomena *amplària d'un interval* o recorregut a la distància que hi ha entre els extrems.

S'anomena *densitat de freqüència absoluta* d'un interval al quocient entre la freqüència absoluta de l'interval i la seua amplària.

Mètode l'arrel: Segons aquest mètode el nombre de classes és igual a l'arrel quadrada del nombre de dades:

Nombre classes = $\sqrt{nombre\ de\ dades}$.

El pas següent és calcular l'amplària dels intervals.

En conseqüència, $l'amplària \approx \dfrac{recorregut}{nombre\ de\ classes}$.

Amb aquesta informació, es pot començar sense dubtes la solució del problema.

Exercici 4

Cal recordar però, que els diferents valors que pot prendre la variable estadística es denoten mitjançant x_i. En aquest cas, ordenant-los de menor a major, $x_1 = 5,\ x_2 = 6,\ x_3 = 7, x_4 = 10,\ x_5 = 11,\ x_5 = 12,\ x_6 = 15$.

S'anomena *freqüència absoluta del valor* x_i al nombre de vegades que apareix repetida l'observació en el recull de dades. Es representa per n_i. La freqüència absoluta del valor x_2 és 2 ($n_2 = 2$), ja que la dada 6 es repeteix dues vegades en el conjunt de les dades de la mostra.

S'anomena *freqüència relativa del valor* x_i al quocient entre la freqüència absoluta d'x_i i el nombre total de dades *n*. Es representa per f_i i evidentment, és la proporció en què es troba el valor x_i dintre del conjunt de dades en tant per u; $f_i = \frac{n_i}{n}$. A l'exemple $f_2 = \frac{n_2}{n} = \frac{2}{20} = 0{,}1$. Per tant, el 10% de les dades són sisos.

És important remarcar que per a calcular freqüències acumulades, les quals anomenarem F_i com a freqüència relativa acumulada i N_i com a freqüència absoluta acumulada, és necessari que les variables a estudiar siguen *ordenables*, és a dir, ha de ser possible establir una relació d'ordre entre les variables. D'una altra manera, no té cap sentit realitzar els esmentats càlculs.

Aquestes definicions permeten resumir les dades. No obstant, la manera més adequada per a sintetitzar les dades és mitjançant la taula de freqüències. Hi apareixen distribuïdes les dades segons les freqüències. Al mateix temps reflecteix tots els conceptes esmentats amb anterioritat.

Exercici 5

a) En els exercicis anteriors ja hem vist que és necessari conèixer la classificació de les variables:

b) La classificació del tipus de variables, com ja es coneix de l'exercici anterior és:
   - Les *variables qualitatives* (*Ordinals o Nominals*).
   - Les *variables quantitatives* (*Discretes o Contínues*).

c) Per a completar la taula de freqüències hem de conèixer:
   – Saber crear la taula de dades contínues. En aquest cas, els intervals ja els tenim, només hem d'afegir-hi la marca de classe.
   – Completar la taula amb les diverses freqüències: n, f, N i F.
   – A més a més, cal conèixer els passos per a crear la taula si no coneixem els interval, però açò és un problema que no tenim en aquest cas.

d) Hem de buscar en la taula el valor en els intervals. En aquest cas, l'interval que té el màxim en 150000.

e) Es demana el percentil 65. Els percentils divideixen a la distribució en 100 parts (99 divisions). P1,..., P99, corresponents a 1%,..., 99%. En aquest cas, el valor corresponent al percentil 30, hi té un 30 % de les dades superiors o iguals.

Exercici 6

Com a primera ajuda recordeu que:

– Cal saber el tipus de variable. En aquest cas variable quantitativa contínua, ja que es mostra un gràfic amb format d'histograma.
– A més a més, recordareu que en l'eix de les ordenades, el que apareix és la freqüència relativa, ni l'absoluta ni cap de les acumulades. Aquesta suposició és perquè el gràfic no està en tot moment augmentant.
– El format de la taula de freqüències tindrà la forma:

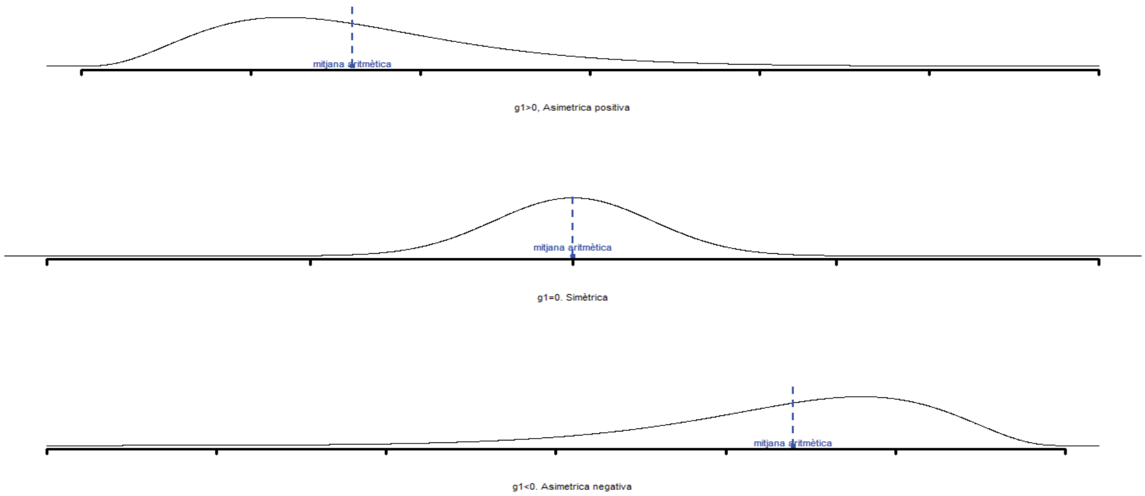
$[L_{i-1}, L_i)$	ni	Ni	fi	Fi

Exercici 7

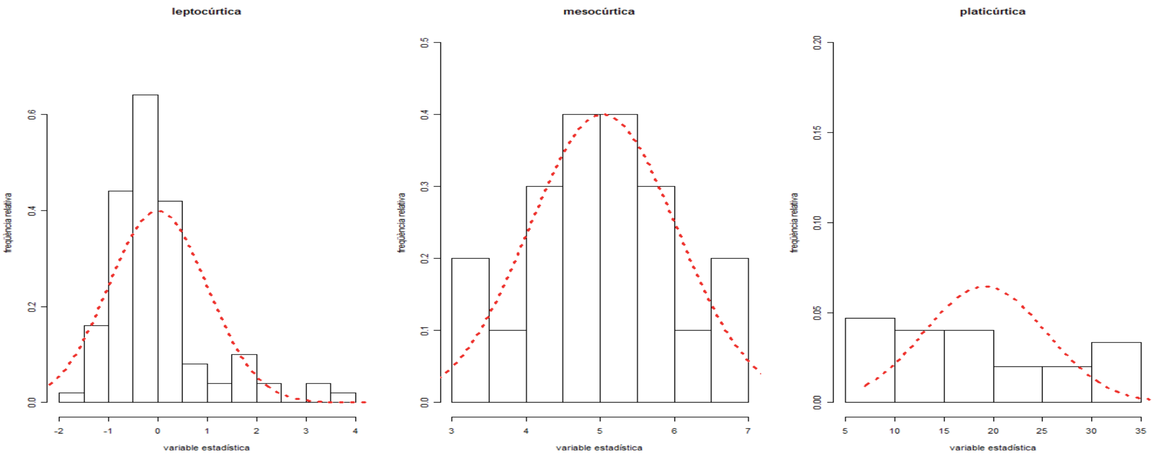
a) Pel que fa al càlcul de la mitjana aritmètica, cal tenir en compte que és una variable contínua i que cal utilitzar la marca de classe en cada cas.

b) Respecte a la forma de les distribucions, hem de treballar els coeficients d'asimetria i curtosis, a més de crear els gràfics per veure la forma de les distribucions. En aquest cas es pot utilitzar el diagrama de barres aplicat a cada empresa per veure la desviació respecte a la distribució normal.

La forma dels gràfics pot ser respecte a la simetria:



Respecte a l'apuntament:



Exercici 8

a) Respecte al que se'ns pregunta en el primer apartat, és el coeficient de variació de cadascuna de les companyies i després realitzar-ne la comparació.

b) En el segon apartat es pregunta l'índex de concentració de Gini. Per a calcular-lo, se seguirà la informació següent:

Se suposa que hi ha una distribució de rendes ($x_i \cdot n_i$) on i pren els valors d'1 fins a k (per exemple, x_i *són els sous* i n_i el nombre de persones que cobren aquest sou) de la qual es formarà una taula amb les columnes següents:

1) Els productes, $x_i \cdot n_i$ indicaran la renda total percebuda pels n_i rendistes de renda individual x_i.

2) Les freqüències absolutes acumulades N_i.

3) Els totals acumulats u_i que es calculen de la forma següent:

$$
\begin{aligned}
u_1 &= x_1 n_1 \\
u_2 &= x_1 n_1 + x_2 n_2 \\
u_3 &= x_1 n_1 + x_2 n_2 + x_3 n_3 \\
u_4 &= x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
u_k &= x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4 + \ldots\ldots\ldots\ldots + x_k n_k
\end{aligned}
$$

Per tant, es pot dir que :

$$u_j = \sum_{i=1}^{j} x_i \cdot n_i \text{ per a qualsevol valor de } j \text{ des d'1 fins a } k.$$

4) La columna total de freqüències acumulades relatives, que s'expressa en tant per cent i que es representa per p_i, vindrà donada per la següent notació:

$$p_i = \frac{N_i}{n}$$

5) La columna de renda acumulada relativa, que s'expressa en tant per cent i que es representa per l'expressió:

$$q_i = \frac{u_i}{u_k}$$

Aleshores ja es pot confeccionar la taula:

x_i	n_i	$x_i n_i$	N_i	u_i	$p_i = \frac{N_i}{n}$	$q_i = \frac{u_i}{u_k}$	pi - qi
x_1	n_1	$x_1 n_1$	N_1	u_1	p_1	q_1	$p_1 - q_1$
x_2	n_2	$x_2 n_2$	N_2	u_2	p_2	q_2	$p_2 - q_2$
...	...	...	...	...	...	...	...
x_k	n_k	$x_k n_k$	N_k	u_k	100	100	0

Exercici 9

a) Les dades en aquest cas són molt importants, ja que s'hi mostren les dades agrupades però en diferent amplitud. Per aquest motiu, s'ha de representar la seua densitat, no directament les dades que se'ns presenten.

b) Per a obtenir les dades de la mitjana aritmètica, cal tenir en compte el mateix element que s'ha comentat amb anterioritat, que són dades agrupades. Si la variable està agrupada en intervals el concepte no canvia. En aquest cas, s'assignen les freqüències a les marques de classe i es procedeix de la mateixa manera que en el cas de les no agrupades.

c) L'estudi de la dispersió està relacionat amb el càlcul de la desviació típica en el cas del treball de variables per separat, però en aquesta ocasió, per comparar-les, s'utilitza el coeficient de variació de Pearson.

d) Una possibilitat és obtenir el gràfic de caixes i bigots.

Ajudes Tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas necessari a l'hora de realitzar els exercicis i problemes, i després de consultar l'ajuda de tipus 1.

Exercici 1

Encara que es coneixe la classificació de les variables, i es té prou informació per a classificar els distints apartats, es poden afegir exemples de cada cas per comparar-los amb els que es demanen:

– *Variables qualitatives nominals*: el sexe o el color.
– *Variables qualitatives ordinals*: estar bé, regular o malalt i també estar ple, mig ple o buit.
– *Variables quantitatives discretes:* nombre de treballadors en una empresa o nombre d'edificis en un carrer.
– *Variables quantitatives contínues:* l'alçada de les persones, les qualificacions numèriques d'un examen o la mesura en centímetres de la fabricació de taules.

Exercici 2

Ja coneixeu quins tipus de gràfics s'han d'utilitzar en cada cas. Ara heu de seguir els passos següents per a fer els gràfics:

– Crear la taula de freqüències corresponent, que en aquest cas, com que és una variable quantitativa discreta no caldrà formar intervals i després fer els gràfics corresponents amb les seues freqüències.

Primerament, crearem la taula de freqüències per a poder crear els gràfics corresponents:

xi	ni	Ni	fi	Fi
0	15	15	0,15	0,15
1	21	36	0,21	0,36
2	21	57	0,21	0,57
3	27	84	0,27	0,84
4	16	100	0,16	1,00
Total	100		1	

a) Respecte a les representacions gràfiques, com que es refereixen a dades discretes, hem d'utilitzar un gràfic que pot ser el de sectors o el de barres. Es representa en l'eix d'abscisses les classes que en aquest cas és el nombre de fills i en l'eix d'ordenades la freqüència corresponent, que pot ser tant l'absoluta com la relativa (acumulada o no).

b) Per a construir el polígon de freqüències amb les freqüències acumulades, s'utilitzaran també les dades de la taula de freqüències i pot ser tant la N com la F.

Exercici 3

El pas següent és, amb el coneixement de creació de les dades agrupades, generar la taula de freqüències agrupada com queda a continuació:

$[L_{i-1}, L_i)$	ni	Ni	fi	Fi
[3,25–3,75)	3	3	0,075	0,075
[3,75–4,25)	8	11	0,2	0,275
[4,25–4,75)	14	25	0,35	0,625
[4,75–5,25)	6	31	0,15	0,775
[5,25–5,75)	4	35	0,1	0,875
[5,75–6,25)	5	40	0,125	1
	N = 40			

Per a cada apartat:

a) Crearem un histograma de forma general.
b) Es crearà un histograma amb quatre classes, sense realitzar la separació general de les dades agrupades que, com ja se sap és l'arrel de les dades.

Exercici 4

El format de la taula serà:

x_i	n_i	N_i	f_i	F_i
Total				

I una forma de saber si les dades es mantenen o varien al llarg del temps seria crear un gràfic de barres, per ser una variable numèrica discreta.

Exercici 5

a) La variable d'estudi és la quantitat d'euros que es gasten les famílies per comprar la segon vivenda. Dada que ens ajudarà en la solució dels propers apartats.

b) El format de la taula serà:

Euros	Marca	Famílies (n_i)	f_i	N_i	F_i

c) Es pot utilitzar la freqüència relativa acumulada i restar a 1 el valor de la F_i anterior.

d) Es demana el percentil 65.

En distribucions agrupades, cal determinar l' interval $[L_{i-1}, L_i)$ en el qual es troba el quantil. Aquest interval es determina seguint exactament els mateixos procediments esmentats a l'apartat anterior; es realitza el mateix que en el cas de dades no agrupades. La diferència radica en què s'obtindrà un interval en lloc d'un valor.

Una vegada es té l' interval $[L_{i-1}, L_i)$, el quantil es calcula:

$$Quantil = L_{i-1} + \frac{\frac{a}{100} \cdot n - N_{i-1}}{n_i} a_i \text{ on,}$$

L_{i-1} Límit inferior de la classe del percentil.
N_{i-1} És la freqüència absoluta acumulada de la classe «anterior» a la classe del percentil.
n_i És la freqüència de la classe del percentil.
a_i És l'amplària de la classe del percentil.

Exercici 6

Tots els apartats següents al de la creació de la taula en depenen, la qual ajuda a calcular cadascun dels estadístics.

a) Ens pregunten: la mitjana, la moda i la mediana.
b) El rang, com ja sabreu és la diferència entre màxim i mínim valor de la variable i el rang interquartílic és la diferència entre el quartil primer i tercer.
c) Ens pregunten el percentil 80, ja que ens parla dels valors més alts i a partir d'aquest, es calcula el percentatge.
d) Es deu aplicar el teorema de Thebyshev a partir de la desviació típica.
e) Aplicació de les propietats de la mitjana, la moda i la mediana, on tots els factors que sumen, resten, multipliquen o dividisquen a la variable afecten. En aquest cas: es multiplicaria per 0,05, pel 5% i se sumaria als tres estadístics 50 euros.
f) El que es demana és comparar la variabilitat o la dispersió de dues mostres diferents. En aquests casos, el més correcte és calcular el coeficient de variació de Pearson, CV. Per aquest motiu, és necessari calcular tant la mitjana com la desviació típica de les variables.

Exercici 7

Per l'apartat *b*), cal conèixer la forma dels coeficients d'asimetria i de curtosi:

D'aquesta manera s'obté el *coeficient d'asimetria de Fisher.*

$$g_1 = \frac{m}{s^3} = \frac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^3 n_i}{n}}{\left(\sqrt{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^2 n_i}{n}}\right)^3}$$

Cal notar que, com la desviació típica és positiva, el signe del coeficient de Fisher serà el mateix que el de *m*. I per tant:

si $g_1 = 0$	*la distribució és simètrica*
si $g_1 > 0$	*la distribució és asimètrica positiva*
si $g_1 < 0$	*la distribució és asimètrica negativa*

Així doncs, quan $g_1<0$, es diu que la distribució presenta asimetria a l'esquerra (o negativa) i llavors, de les dues branques de la corba que separa l'ordenada que passa per la mitjana aritmètica, la de l'esquerra és més llarga que la de la dreta. L'oposat ocorre si $g_1>0$.

De la mateixa manera que en el cas de l'estudi de la asimetria, hi ha un coeficient que permet classificar les dades segons la curtosi. En aquest cas, el coeficient no és tan intuïtiu, i per això únicament se'n donarà la definició i la seua interpretació. Com en el cas de l'altra mesura de forma, aquest indicador tampoc no té dimensió.

$$g_2 = \frac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^4 n_i}{n}}{\left(\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^2 n_i}{n}\right)^2} - 3$$

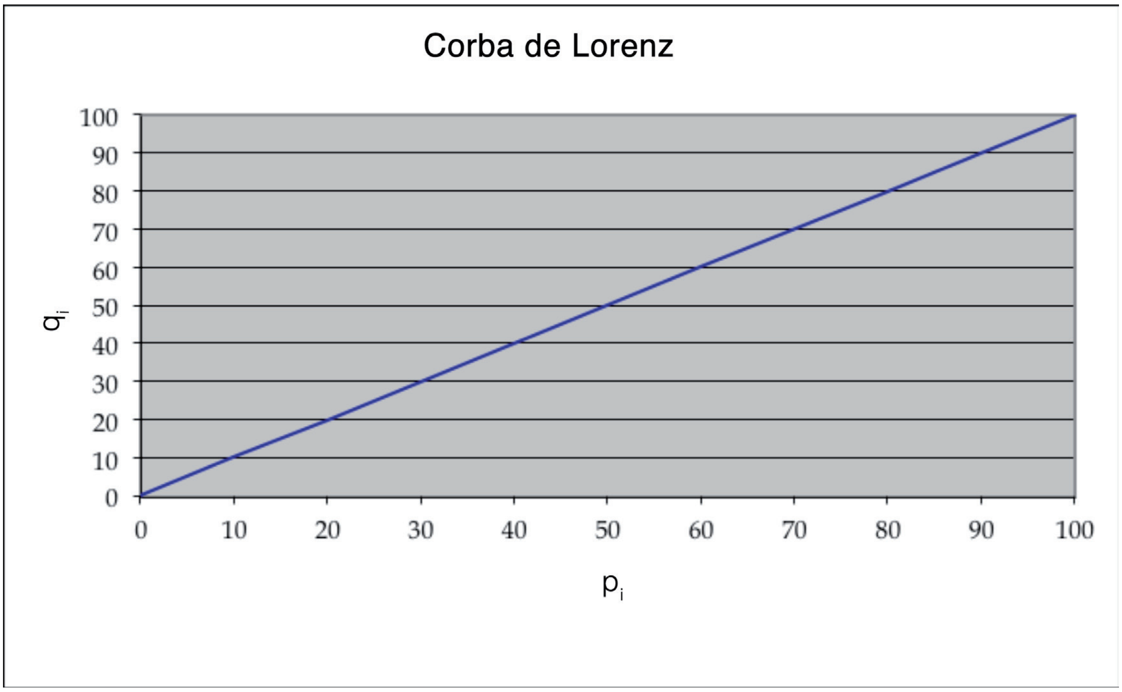
La idea de l'apuntament d'una distribució de dades surt de la comparació de la freqüència dels valors centrals d'una distribució amb la freqüència dels valors centrals en un model teòric normal que tinga la mateixa mitjana i la mateixa desviació típica que la distribució que s'hi estudia.

Com que en un model normal es compleix que $\dfrac{\dfrac{\sum_{i=1}^{k}(x_i - \overline{X})^4 n_i}{n}}{s^4} = 3$, llavors: una distribució serà:

mesocúrtica (normal)	si $g_2 = 0$
leptocúrtica	si $g_2 > 0$
platicúrtica	si $g_2 < 0$

Exercici 8

Com a ajuda final, es pot comentar el tipus de gràfic que s'hi ha de representar. És la corba de Lorenz. La forma d'aquest gràfic és la següent:



S'hi ha de representar els valors de q_i front als valors de p_i.

La línia central és la línia d'equitat de les dades, que ens marcarà com de separat estarà de la concentració.

Exercici 9

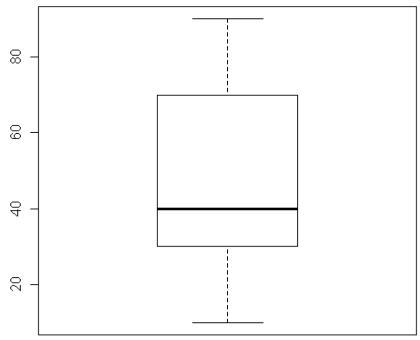
a) L'obtenció de les densitats serà el quocient, en cada cas, de la freqüència absoluta entre l'amplitud de l'interval.

b) Un diagrama de caixes i bigot (conegut també com *Box and whisker Plot* en anglès), és una representació gràfica de les dades que permet determinar amb molta facilitat i d'una manera visual la tendència central, la variabilitat, la asimetria i l'existència de valors anòmals d'un conjunt d'observacions. D'alguna manera, es pot dir que és un dels gràfics que més i millor resumeixen els conjunts de dades.

El diagrama de caixes empra el resum dels 5 nombres: la menor observació, la major observació, el primer quartil, la mediana i el tercer quartil. Aquests 5 nombres permeten construir la versió més simple del Box-Plot, el qual està format per:

*Una caixa (*box*) central* que representa les observacions compreses entre el primer i el tercer quartil. Els dos extrems de la caixa són els quartils, i una línia interior i vertical que parteix la caixa en dos parts, correspon a la mediana. És obvi doncs, que la caixa comprén el 50% de les observacions.

*Bigots (*Whiskers*)*: El gràfic es completa en aquesta versió del Box-Plot, amb dues línies a ambdós costats de la caixa que uneixen el primer quartil amb la menor observació, i el tercer quartil amb l'observació major.



Solucions completes

Exercici 1

Classifiqueu les variables següents, justificant el perquè de l'elecció:

a) Color dels cotxes.
b) Marques d'ordinadors.
c) Longitud de carreteres en metres.
d) Nivell d'estudis.
e) Nombre de fills d'una família.
f) Nombre d'alumnes d'estadística en una carrera.
g) Metres d'altitud de les muntanyes.
h) Professions de les persones.
i) Sou mensual dels treballadors de les empreses del sector ceràmic.

Solució

a) És una variable qualitativa nominal: color A, color B, color C, etc.
b) És una variable qualitativa nominal: marca X, marca Y, marca Z, etc.
c) És una variable quantitativa contínua: 1.93, 1.935, 1.76, 1.67, etc.
d) És una variable qualitativa ordinal: sense estudis, elementals, etc.
e) És una variable quantitativa discreta: 0, 1, 2, 3, etc.
f) És una variable quantitativa discreta: 0, 1, 12, 3033, 5004, etc.
g) És una variable quantitativa contínua: 36.1, 36.51, 36.512, 36.78, 37.1, 39.12, etc.
h) És una variable qualitativa nominal: metge, professor, pallasso, etc.
i) És una variable quantitativa contínua: 1200.50, 1165.43, 1500.23, etc.

Exercici 2

Actualment, s'estudia en les distintes comunitats autònomes el nombre de fills per família per a analitzar la natalitat. Un dels treballadors que està fent les enquestes, arreplega les dades del seu barri on hi ha 100 famílies. Hi ha obtingut les dades següents:

1	3	3	0	4	3	1	4	0	0
2	1	0	3	1	2	1	4	1	2
3	3	4	2	0	4	3	0	2	3
1	3	4	2	2	4	4	4	2	1
4	2	1	1	0	1	1	2	3	0

3	3	3	1	1	3	3	0	2	3
4	3	0	3	1	2	2	1	2	3
3	2	1	3	1	3	4	4	4	1
3	0	3	1	0	4	3	2	3	2
1	2	0	2	0	0	2	2	3	4

a) Construïu el gràfic que considereu més adequat per a aquestes dades.
b) Construïu el polígon de freqüències amb les freqüències acumulades.

Solució

El primer pas serà saber quin tipus de variable és, ja que aquest element afectarà l'elecció tant del tipus de taula de freqüències com del tipus de gràfic.

Queda clar que és una variable numèrica. Per tant, pot ser contínua o discreta. En aquest cas, com que les dades fan referència al nombre de fills serà quantitativa discreta.

Amb aquestes informacions, es pot passar a resoldre el problema.

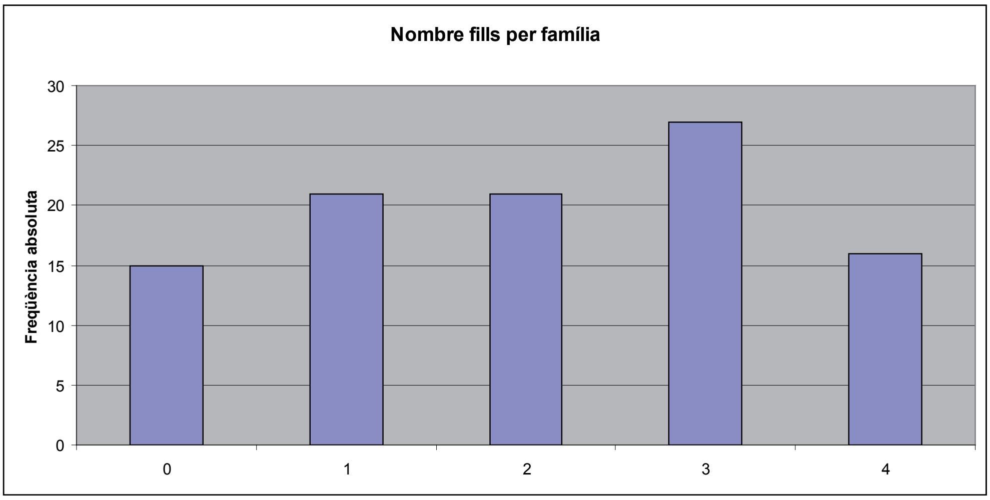
Primerament, crearem la taula de freqüències per a poder crear els gràfics corresponents:

xi	ni	Ni	fi	Fi
0	15	15	0,15	0,15
1	21	36	0,21	0,36
2	21	57	0,21	0,57
3	27	84	0,27	0,84
4	16	100	0,16	1,00
Total	100		1	

a) Respecte a les representacions gràfiques, com que es refereix a dades discretes, hem d'utilitzar un gràfic com ara el de sectors o el de barres. En cap cas usarem l'histograma ja que s'utilitzarà per a les dades contínues.

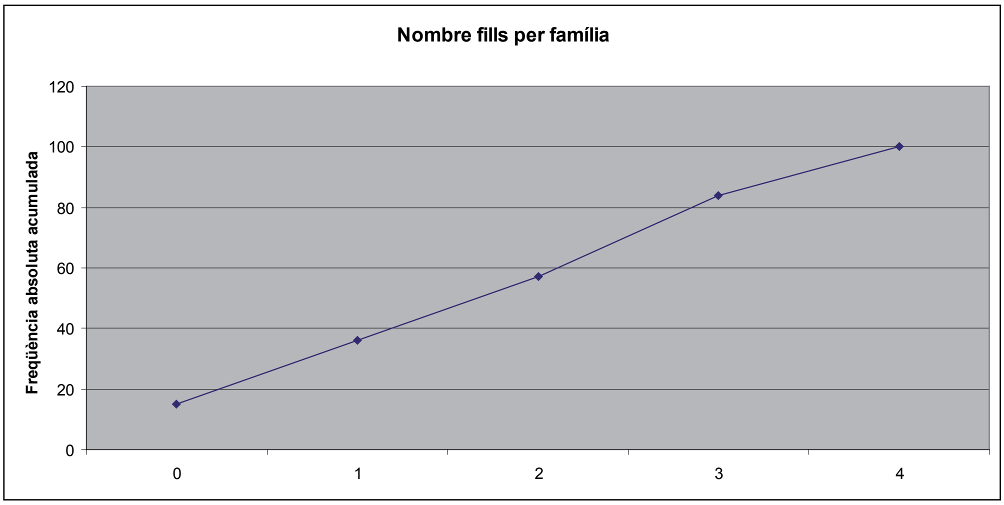
Es representa en l'eix d'abscisses les classes que en aquest cas és el nombre de fills i en l'eix d'ordenades la freqüència corresponent, que pot ser tant l'absoluta com la relativa (acumulada o no).

El resultat de representar la freqüència absoluta en un diagrama de barres, és el següent:



b) Per a construir el polígon de freqüències amb les freqüències acumulades, s'utilitzaran també les dades de la taula de freqüències i pot ser tant la N com la F.

Com es pot veure a continuació, el que s'utilitza com a resolució és la freqüència absoluta acumulada en l'eix d'ordenades.



Exercici 3

Els sous, en milers d'euros mensuals, de 40 empresaris del sector de la construcció de l'any 2007 són:

3,9	4,7	3,7	5,6	4,3	4,9	5,0	6,1	5,1	4,5
5,3	3,9	4,3	5,0	6,0	4,7	5,1	4,2	4,4	5,8
3,3	4,3	4,1	5,8	4,4	4,8	6,1	4,3	5,3	4,5
4,0	5,4	3,9	4,7	3,3	4,5	4,7	4,2	4,5	4,8

Es vol estudiar si realment són prou alts i quina és la seua distribució. Per a aconseguir-ho:

a) Representeu gràficament la informació arreplegada.

b) Creeu la mateixa representació en 4 classes.

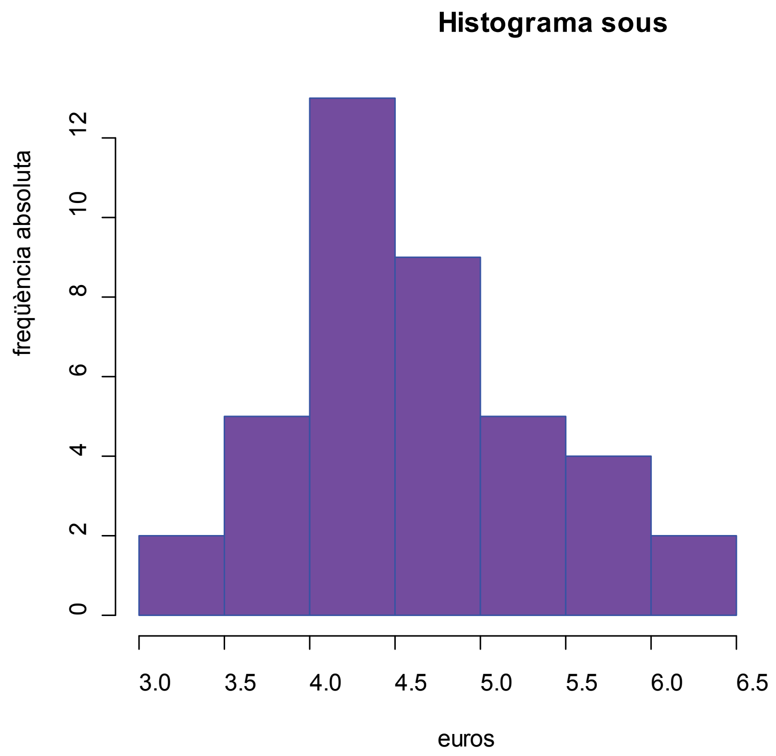
Solució

El primer pas serà saber quin tipus de variable és, ja que aquest element afectarà a l'elecció tant del tipus de taula de freqüències com a l'elecció del tipus de gràfic. Queda clar que és una variable numèrica contínua.

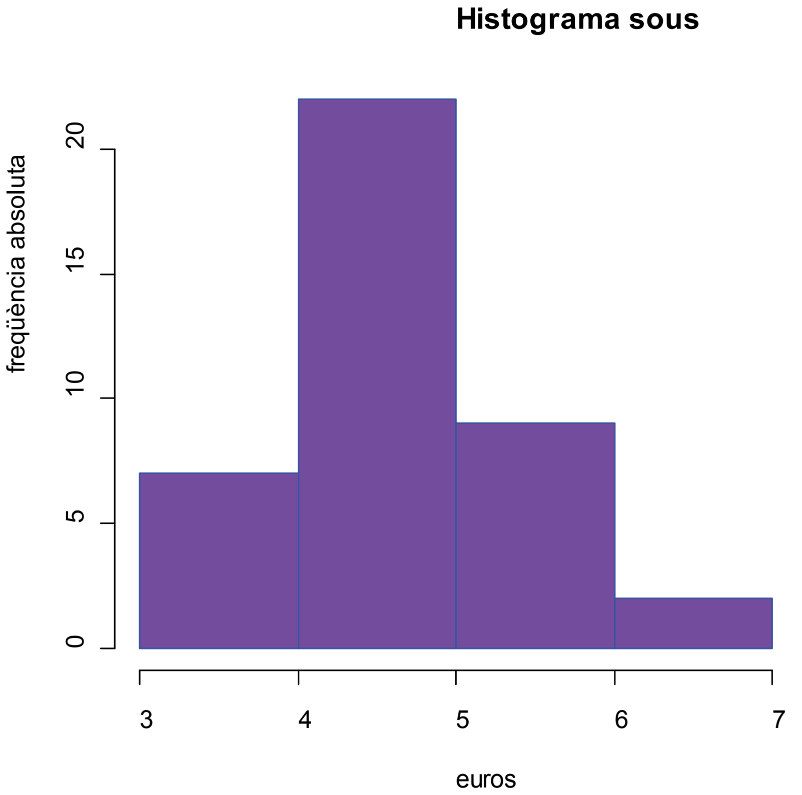
El pas següent és, amb el coneixement de creació de les dades agrupades, fer la taula de freqüències agrupada:

$[L_{i-1}, L_i)$	ni	Ni	fi	Fi
[3,25–3,75)	3	3	0,075	0,075
[3,75–4,25)	8	11	0,2	0,275
[4,25–4,75)	14	25	0,35	0,625
[4,75–5,25)	6	31	0,15	0,775
[5,25–5,75)	4	35	0,1	0,875
[5,75–6,25)	5	40	0,125	1
	N = 40			

a) Un possible resultat, serà l'histograma següent:



b) Com que ens demanen quatre classes, l'histograma passarà a ser el següent:



Exercici 4

El recull de 20 dades corresponents al nombre de telefonades enregistrades en una empresa durant els dies de preparació de material per a una fira de mostres durant el període de 9 a 12 hores.

15 ,5, 10, 5, 5, 6, 5, 6, 5, 6, 7, 10, 10, 12, 11, 11, 12, 15, 12, 15

Es vol estudiar si realment hi ha variació al llarg dels dies de les telefonades que s'hi reben. Per aquest motiu es demana confeccionar una taula de freqüències que reculla aquesta informació.

Solució:

Cal recordar però, que els diferents valors que pot prendre la variable estadística es denoten mitjançant x_i. En aquest cas, ordenant-los de menor a major, $x_1 = 5,\ x_2 = 6,\ x_3 = 7, x_4 = 10,\ x_5 = 11,\ x_5 = 12,\ x_6 = 15$.

S'anomena *freqüència absoluta del valor* x_i al nombre de vegades que apareix repetida l'observació en el recull de dades. Es representa per n_i. La freqüència absoluta del valor x_2 és 2 ($n_2 = 2$), així doncs, la dada 6 es repeteix dues vegades en el conjunt de les dades de la mostra.

S'anomena *freqüència relativa del valor* x_i al quocient entre la freqüència absoluta d'x_i i el nombre total de dades *n*. Es representa per f_i i evidentment, és la proporció en què es troba el valor x_i dintre del conjunt de dades en tant per u; $f_i = \frac{n_i}{n}$. A l'exemple $f_2 = \frac{n_2}{n} = \frac{2}{20} = 0{,}1$. Per tant, el 10% de les dades són sisos.

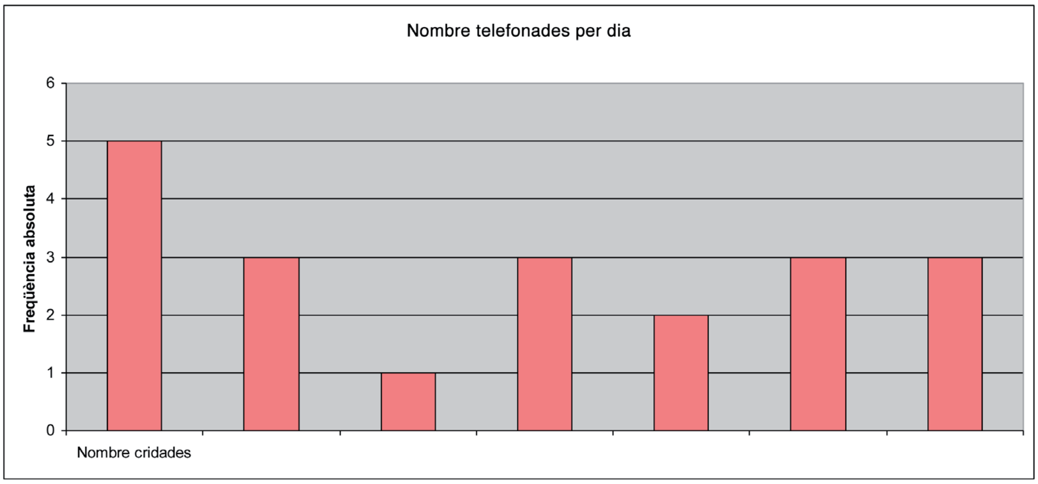
És important remarcar que per a calcular freqüències acumulades, a les quals anomenarem F_i com a freqüència relativa acumulada i N_i com a freqüència absoluta acumulada, és necessari que les variables a estudiar siguen *ordenables*, és a dir, ha de ser possible establir una relació d'ordre entre les variables. D'altra manera, no té cap sentit realitzar els esmentats càlculs.

Aquestes definicions permeten resumir les dades. No obstant, la manera més adequada per a sintetitzar les dades és mitjançant el que es denomina taula de freqüències. Hi apareixen distribuïdes les dades segons les freqüències. Al mateix temps hi reflecteix tots els conceptes esmentats amb anterioritat.

Amb totes aquestes dades, el resultat de la taula serà el següent:

x_i	n_i	N_i	f_i	F_i
5	5	5	0,25	0,25
6	3	8	0,15	0,4
7	1	9	0,05	0,45
10	3	12	0,15	0,6
11	2	14	0,1	0,7
12	3	17	0,15	0,85
15	3	20	0,15	1
Total	20		1	

Si a més a més, representem la freqüència absoluta, podem veure que realment no augmenta el nombre de telefonades, i que es manté prou estable.



Exercici 5

Una empresa està fent l'estudi dels diners que es gasta la gent per a comprar una segon casa com a complement de la primera vivenda. Recolliu les dades dels euros i en nombre de famílies que han comprat aquest tipus de vivenda. A continuació es poden veure les dades:

Euros	Famílies
0-50000	2145
50000-75000	1520
75000-100000	840
100000-115000	955
115000-135000	1110
135000-140000	2342
140000-150000	610
150000-200000	328
>200000	150

Es demana:

a) De quin tipus de variable és l'objecte d'estudi?
b) Mostreu en forma de taula de freqüències el conjunt de les dades arreplegades.
c) Quin percentatge de famílies es gasten més de 150000 euros?
d) El 65% de famílies que menys es gasta, quina quantitat de diners desemborsa com a màxim?

Solució

a) La variable d'estudi és la quantitat d'euros que es gasten les famílies per a comprar la segon vivenda.

b) Per a completar la taula de freqüències hem de conèixer:
   – El tipus de variable que es treballa. En aquest cas és una variable quantitativa contínua.
   – Saber crear la taula de dades contínues. En aquest cas, els intervals ja els tenim, només hem d'afegir-hi la marca de classe.
   – Completar la taula amb les diverses freqüències: n, f, N i F.

La taula que se'ns demana serà:

Euros	Marca	Famílies (n_i)	f_i	N_i	F_i
0-50000	25000	2145	0,2145	2145	0,2145
50000-75000	62500	1520	0,152	3665	0,3665
75000-100000	87500	840	0,084	4505	0,4505
100000-115000	107500	955	0,0955	5460	0,546
115000-135000	125000	1110	0,111	6570	0,657
135000-140000	137500	2342	0,2342	8912	0,8912
140000-150000	145000	610	0,061	9522	0,9522
150000-200000	175000	328	0,0328	9850	0,985
>200000	200000	150	0,015	10000	1

c) Amb l'ajuda de la taula, i amb les dades dels intervals, podem veure quins són els casos superiors a 150000. En aquest cas, per exemple, podem utilitzar la freqüència relativa acumulada:

$$1 – 0,9522 = 0,0488, \text{ que serà un } 4.88\%.$$

d) Es demana el percentil 65.

Els percentils divideixen la distribució en 100 parts (99 divisions). P1,..., P99, corresponents a 1%,..., 99%. En aquest cas, el valor corresponent al percentil 30, té un 30% de les dades superiors o iguals a ell.

En distribucions agrupades, cal determinar l'interval $[L_{i-1}, L_i)$ en el qual es troba el quantil. Aquest interval es determina seguint exactament els mateixos procediments esmentats a l'apartat anterior; s'hi realitza el mateix que en el cas de dades no agrupades. La diferència radica en què s'obtindrà un interval en lloc d'un valor.

Una vegada es té l' interval $[L_{i-1}, L_i)$, el quantil es calcula:

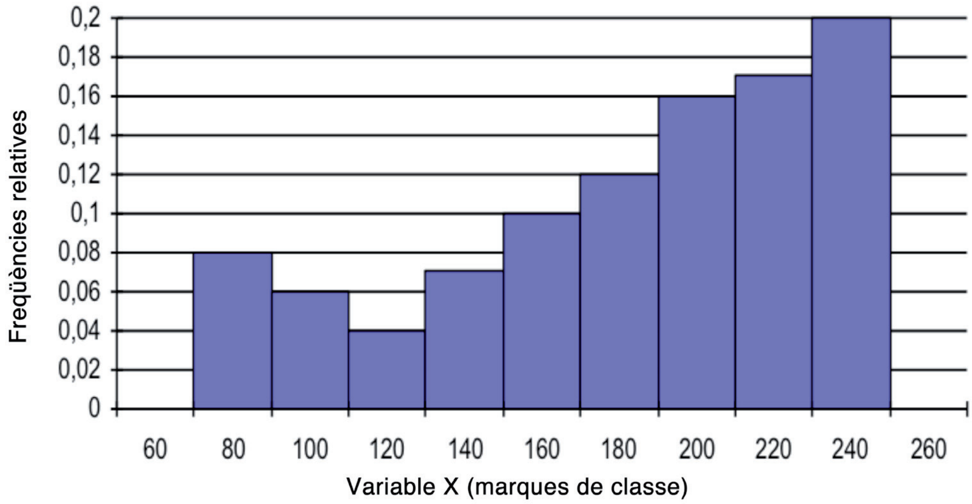
$$Quantil = L_{i-1} + \frac{\frac{a}{100} \cdot n - N_{i-1}}{n_i} a_i \text{ on,}$$

L_{i-1} Límit inferior de la classe del percentil.
N_{i-1} És la freqüència absoluta acumulada de la classe «anterior» a la classe del percentil.
n_i És la freqüència de la classe del percentil.
a_i És l'amplària de la classe del percentil.

En aquest cas: el valor del percentil 65 serà: 133738 euros.

Exercici 6

En l'histograma següent, es representa la distribució dels diners que durant l'últim mes s'han gastat els treballadors d'una empresa en dietes:



Determineu, sabent que hi ha 200 treballadors:

a) La taula de freqüències que mostra les dades que tenim.
b) La quantitat mitjana que s'han gastat, la més freqüent i la quantitat que tenien com a màxim, el 50% dels treballadors que menys cobraven.
c) Calculeu i interpreteu el rang de la distribució així com el rang interquartílic.
d) Calculeu el mínim del 20% dels empleats amb major quantitat de dietes. Quin percentatge del total de l'empresa correspon a aquest grup?
e) L'interval centrat en la quantitat mitjana en què es troben el 75% de les dades. És, doncs, el sou mitjà molt representatiu del conjunt de les dietes?

f) El mes següent, l'empresa decidí augmentar les dietes de tots els treballadors un 5%. A més a més, els hi donà una prima de 50 euros en concepte de productivitat. Calculeu el salari mitjà, el salari més freqüent i el salari que tenien com a màxim, el 50% dels treballadors que menys cobren el mes següent.

g) De les dietes d'una altra empresa, que pertany al mateix sector, se sap que la mitjana aritmètica dels seus treballadors és de 120 euros, amb una variància de 2,5 euros. Quina empresa té una dieta mitjana més representativa? Raoneu-ne la resposta.

Solució

a) Taula de freqüències de dades agrupades a partir d'un gràfic:

$[L_{i-1}, L_i)$	ni	Ni	fi	Fi
[70-90)	16	16	0,08	0,08
[90-110)	12	28	0,06	0,14
[110-130)	8	36	0,04	0,18
[130-150)	14	50	0,07	0,25
[150-170)	20	70	0,10	0,35
[170-190)	24	94	0,12	0,47
[190-210)	32	126	0,16	0,63
[210-230)	34	160	0,17	0,80
[230-250)	40	200	0,20	1
	N = 200		1	

b) $\overline{X}$ =182 / Mo = 240 / Me = 193,750. Tots els estadístics en euros.

c) Rang = 250 – 70 = 180 . Per a determinar el rang interquartílic, s'ha de calcular abans el primer i el tercer quartil: C_3 = 224,11 C_1 = 150   Ri = 74,11

d) El sou mínim és el P_{80} = 230 . La proporció = 26,373%.

e) L'interval es trau aplicant el teorema de Thebyshev: [131,681, 232,319]; ja que la desviació típica és de 50,319 euros.

f) $\overline{X}$ =241 / Mo = 302 / Me = 253,43. Tots els estadístics en milers d'euros.

g) S'ha de calcular el coeficient de variació de totes dues observacions. En la primera empresa, el coeficient de variació és: CV = $\frac{S}{\overline{X}} = \frac{50,319}{182} = 0,276$ i en el segon cas: CV=$\frac{\sqrt{2,5}}{120} = 0,013$.

Per tant, la mitjana aritmètica dels sous de la segon empresa és més representativa que la de la primera.

Exercici 7

Es vol llançar al mercat un nou producte ceràmic i l'empresa que el crea estudia el temps de publicitat, en segons, que unes altres empreses han utilitzat per a promocionar un producte similar. A continuació, es pot veure per a cada empresa la duració i els anuncis realitzats:

Empresa 1

Duració	Nombre d'anuncis
0-20	3
20-25	17
25-30	13
30-40	9
40-60	8

Empresa 2

Duració	Nombre d'anuncis
0-20	1
20-25	5
25-30	13
30-40	5
40-60	2

Empresa 3

Duració	Nombre d'anuncis
0-20	4
20-25	6
25-30	7
30-40	5
40-60	3

Empresa 4

Duració	Nombre d'anuncis
0-20	3
20-25	17
25-30	13
30-40	9
40-60	8

Per a fer l'estudi, calculeu:

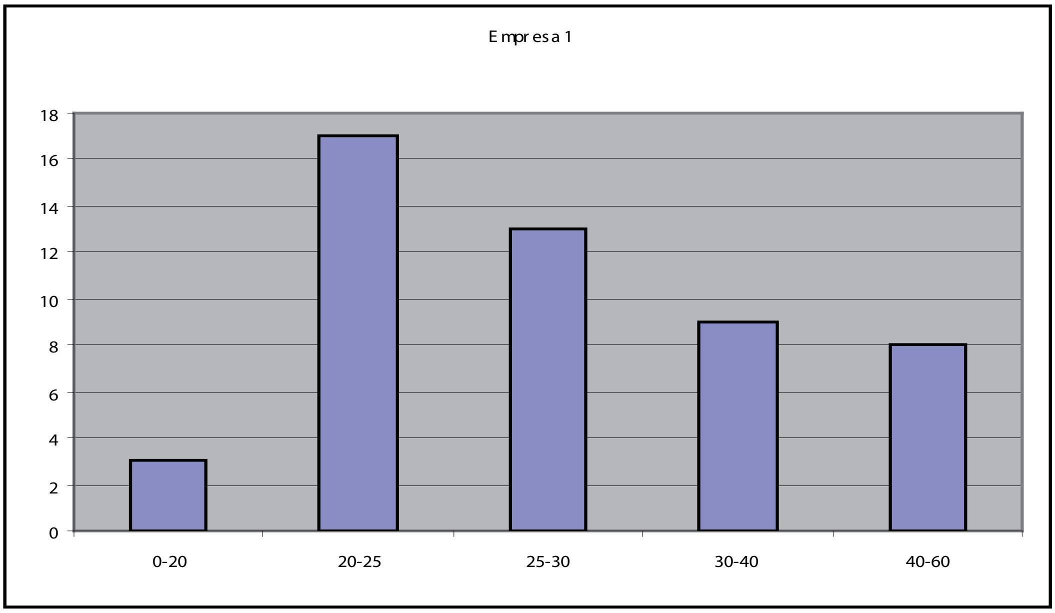
a) La durada mitjana de cada empresa.
b) Tenen totes les distribucions la mateixa forma? Comenteu-ne el resultat.

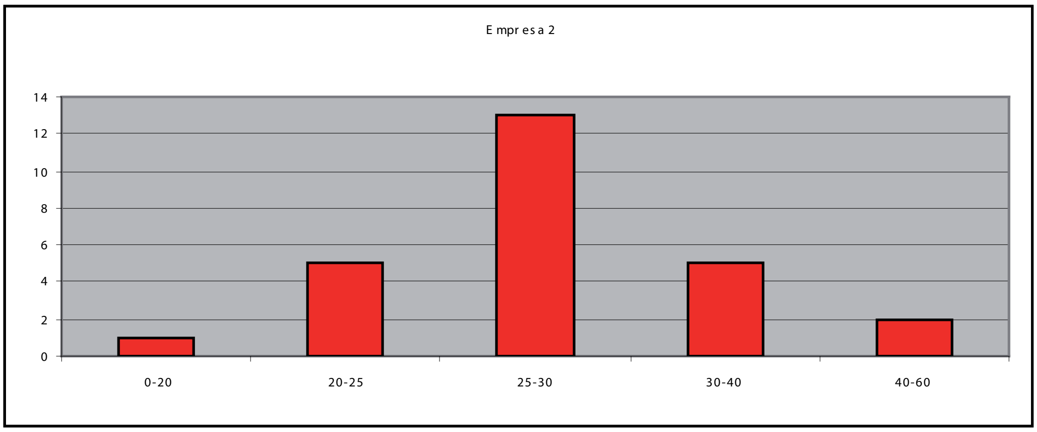
Solució

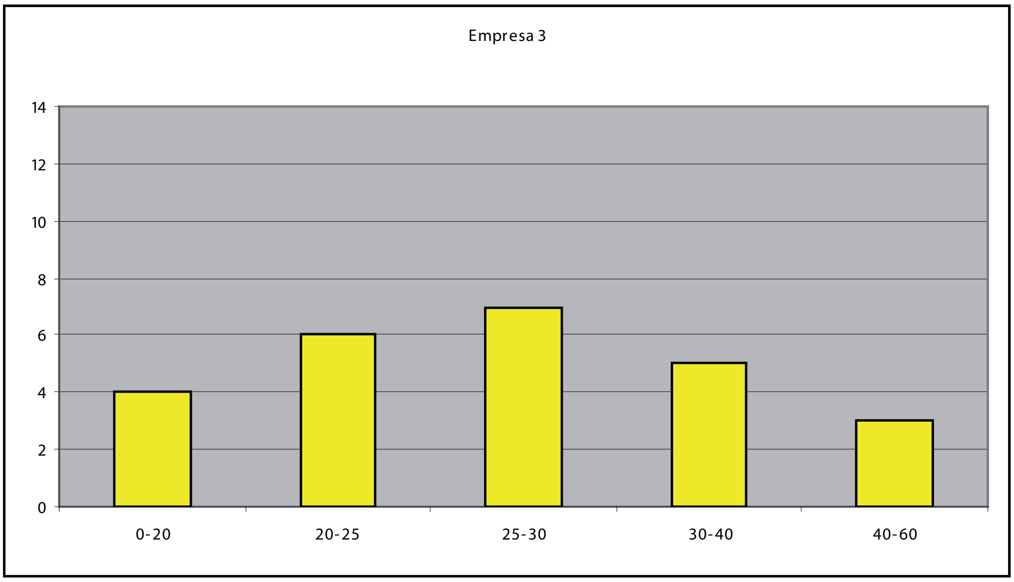
a) La mitjana aritmètica per a cada cas serà:

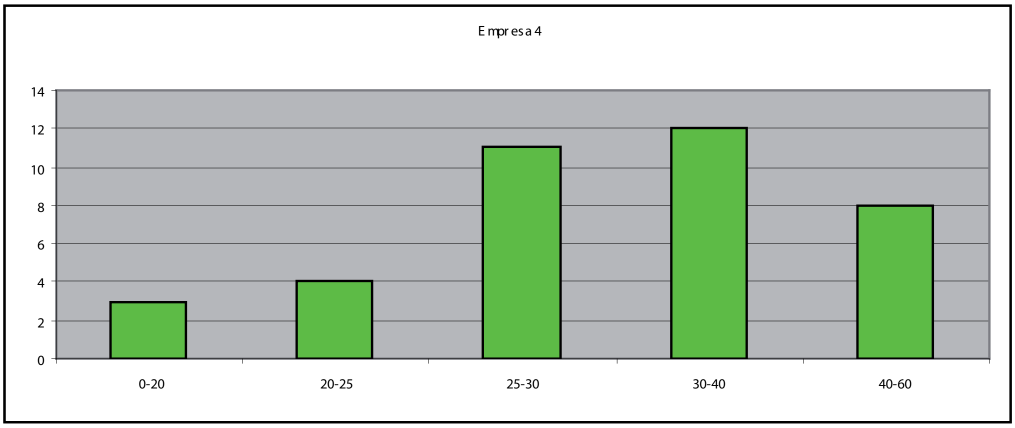
Empresa 1: 29,70 segons.
Empresa 2: 29,04 segons.
Empresa 3: 27,70 segons.
Empresa 4: 32,70 segons.

b) Ho representem en forma de diagrama de barres o histograma de barres per a veure la forma de la distribució:









A més a més, podem calcular els valors dels coeficients d'asimetria i de curtosi en cada cas, per veure clarament que:

– L’empresa 1 es asimètrica a dretes.
– L’empresa 2 es leptocúrtica.
– L’empresa 3 es platicúrtica.
– L’empresa 4 es asimètrica a esquerres.

Asimetria (g1)	0,0506	1,4646	0,0000	-0,1231
Curtosi (g2)	-0,1875	2,4434	-1,2000	-2,7111

Exercici 8

Dues companyies de venda de cotxes tenen maneres diferents de pagar als seus treballadors. La companyia A ho fa mitjançant un sou fix mensual i la companyia B mitjançant un percentatge sobre les vendes efectuades. La distribució dels salaris per categories és la següent:

COMPANYIA A	
Sou (centenes d’euro)	Nombre de treballadors
26	10
39	10
52	40
247	20
260	10
273	10

COMPANYIA B	
Sou (centenes d’euro)	Nombre de treballadors
4	10
5	10
6	40
7	20
26	10
27	10

a) Basant-se únicament en les observacions, en quina companyia el sou mitjà fluctua menys o té els repartiments més equitatius? Justifiqueu el resultat mitjançant l’anàlisi estadística del repartiment.

b) En quina de les dues companyies el sou és més homogeni o concentrat? S’ha d’obtenir el resultat també de forma gràfica.

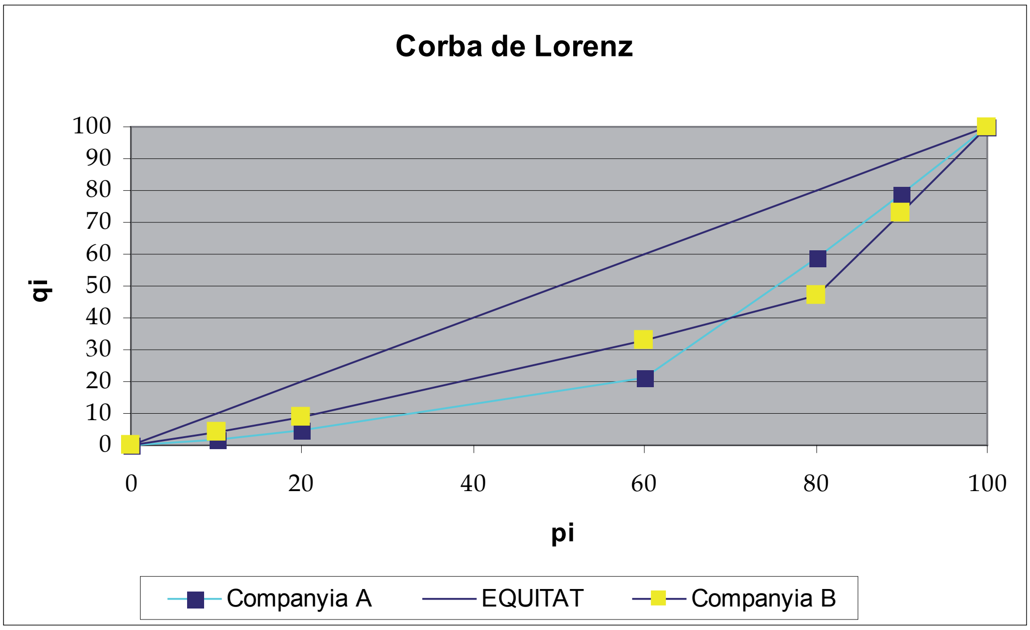
Solució

a) El sou mitjà de la companyia A és de 130 i el coeficient de variació és de 83,2.
El sou mitjà de la companyia B és de 10 i el coeficient de variació és de 6,88.

És a dir, en la companyia A el sou mitjà és menys representatiu de les dades.

b) Totes dues distribucions de dades tenen el mateix índex de Gini: 0,361538. Per tant, en les dues la concentració és igual.

A continuació es pot veure la representació de la corba de Lorez per als dos casos:



Com es pot observar en el gràfic, les dues corbes de Lorenz es creuen, i per això, malgrat tenir distribucions diferents la concentració és la mateixa.

Exercici 9

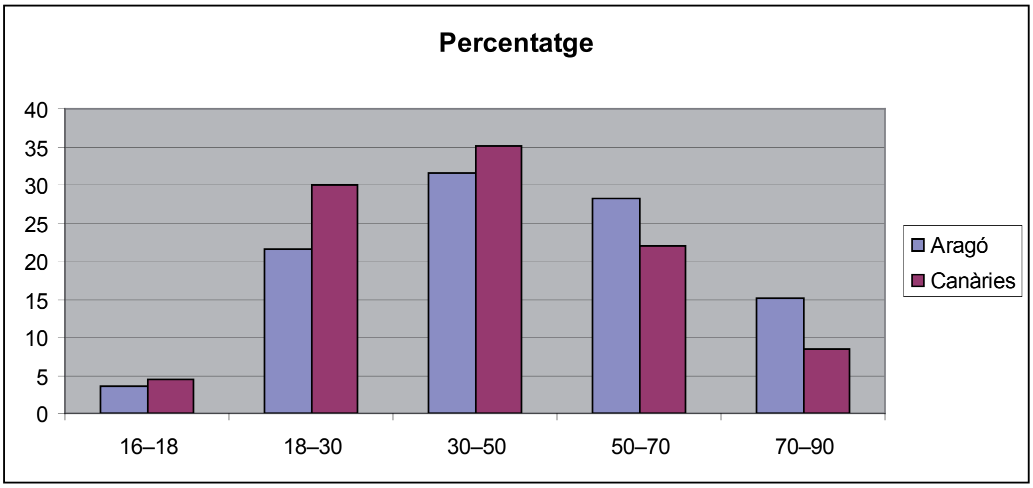
La distribució d'edats del Cens Electoral de Residents a 1 de gener de 1999 per a les comunitats autònomes d'Aragó i Canàries, en tants per cent és la següent:

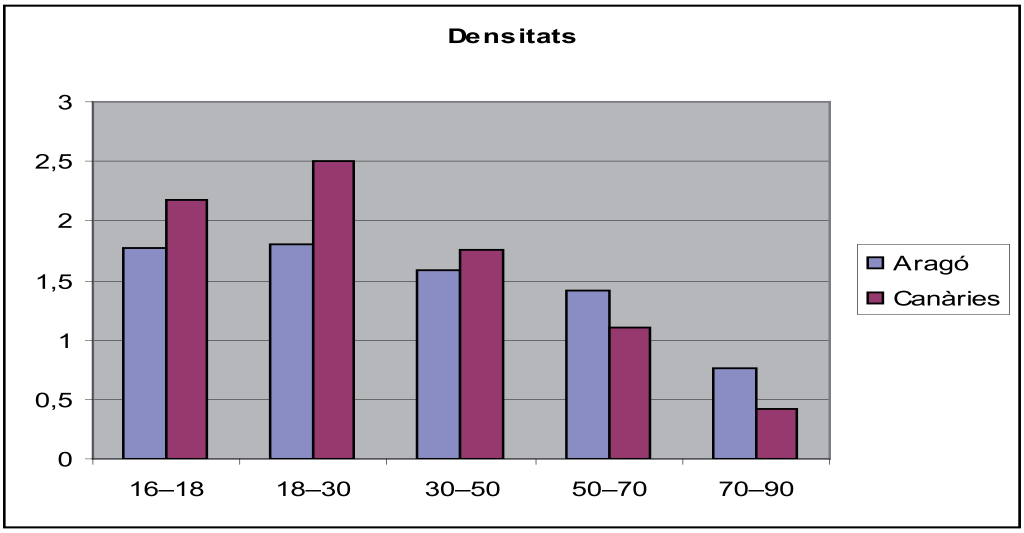
Edats	Aragó	Canàries
16-18	3,55	4,35
18-30	21,56	29,99
30-50	31,63	35,21
50-70	28,14	21,97
70-90	15,12	8,48

a) Representeu sobre els mateixos eixos de coordenades les dades de la distribució de l'edat per a les dues comunitats autònomes (empreu distint traç o distints colors). Quines conclusions obteniu a la vista del gràfic?
b) Calculeu l'edat mitjana per a les dues comunitats. Compareu-les. Què indiquen aquests resultats?
c) En quina comunitat les observacions són més disperses?
d) Si les dades d'edats foren: Aragó: 10, 10, 10, 10, 20, 30, 40, 30, 40, 50, 60, 40, 40, 40, 60, 70, 80, 70, 80, 90, 70, 50, 40, 90. Canàries: 20, 30, 40, 40, 140, 50, 40, 30, 40, 30, 50, 60, 40, 30, 30, 40, 30, 40, 30, 40, 30, 50, 60, 70. Obteniu un gràfic que mostre la dispersió de les dades.

Solució

a) Es representen els dos conjunts de dades tenint en compte que els intervals no tenen la mateixa amplària, i per tant, s'han de calcular les densitats. Podem veure la diferència tal com apareixen les dades i representant la densitat, que serà el correcte:





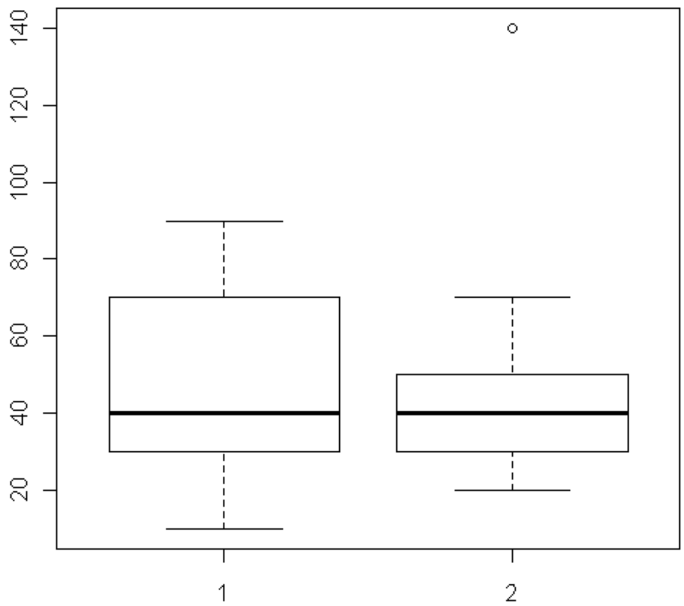
Clarament tenim diferències utilitzant la densitat i que augmenta l'edat en la comunitat d'Aragó respecte a la de Canàries de gent de major edat.

b) L'edat mitjana d'Aragó és de 47,41 anys i la de Canàries de 41,98. Pels resultats, es podria dir que si les mitjanes foren representatives de les dades, la població d'Aragó està una mica més envellida que la de Canàries.

c) Per a saber-ho, cal determinar els coeficients de variació en tots dos conjunts d'observacions: CV Aragó = 40,75%, CV Canàries = 42,56%.

En conseqüència, a les illes Canàries les observacions són més disperses. No obstant, com els coeficients de variació de totes dues comunitats són molt alts, les mitjanes aritmètiques no serien gaire representatives del conjunt de dades en cap cas.

d) El que hem d'obtenir és el gràfic de caixes i bigots per a les dues variables, sent 1 = Aragò i 2 = Canàries:



La dispersió és major a Aragó que a les Canàries.

UNITAT 2

Estadística descriptiva bivariant

Introducció teòrica

Normalment, en qualsevol investigació no s'estudia una única variable dels individus que formen la mostra (referències bibliogràfiques 1, 5, 9, 12 i 16), sinó que en moltes ocasions en són més. Així, si es vol estudiar el rendiment dels treballadors d'una empresa, de cada treballador pot ser útil conéixer l'edat, el sou, el nivell d'estudis, les hores que treballa, el nombre de persones que té al seu càrrec, etc. És a dir, per a cada individu de la mostra s'obté un vector o registre en què cada component és el valor d'una de les variables subjectes a estudi; a l'exemple que s'està considerant un vector associat a un individu:

(35 anys, 24500 €, diplomat, 47 hores setmanals, 2 persones al seu càrrec...).

Aquest fet origina que l'investigador es plantege, a més a més de l'estudi individualitzat de cadascuna de les variables, l'estudi conjunt de totes o d'algunes elles. D'aquesta manera és possible conèixer si existix algun tipus de relació funcional o estadística entre les variables. Així, les observacions poden manifestar que aquelles persones amb más titulació tinguen més persones al seu càrrec, o que a mesura que va augmentant l'edat dels treballadors també ho fa el sou. A més, si aquesta relació existix pot ser que es puga trobar una «fórmula matemàtica» que relacione formalment les variables.

D'altra banda, la nomenclatura canvia si s'estudien conjuntament diferents variables. Així, si es realitza l'estudi de dues variables es diu que es treballa amb variables bidimensionals, si en són tres, variables tridimensionals i si en són més de tres, variables pluridimensionals.

Distribucions estadístiques bidimensionals: taules i gràfics

Quan es volen estudiar dues característiques observables sobre una mateixa mostra o població, cada una de les variables que constitueix la variable bidimensional (X,Y) es denomina component o variable marginal, i poden ser tant un atribut, com una variable quantitativa. En qualsevol cas, al realitzar-se el treball de recollida de dades s'obté un conjunt de parells ordenats del tipus:

$$\{(x_1, y_1), (x_1, y_1),..., (x_2, y_1), (x_2, y_1),..., (x_i, y_j),..., (x_i, y_j),..., (x_h, y_k),..., (x_h, y_k)\}$$

Per exemple, si es considerara X la variable dies d'estudi per a un examen d'Estadística i Y la nota obtinguda per a un conjunt d'estudiants, les dades recollides serien del tipus:

$$\{(5,3) ; (6,5); (5,3) ; (6,5); (5,7)\}$$

En les dades, cada observació es repeteix un nombre determinat de vegades. Així, una primera manera de representar el conjunt de dades és mitjançant la terna següent: $\{(x_i, y_i), n_{i,j}\}$ en què:

x_i representa els valors de la variable X.
y_i representa els valors de la variable Y.
$n_{i,j}$ és el nombre de vegades que es repeteix la dada (x_i, y_j), és a dir, la seua freqüència absoluta.

Seguint amb l'exemple tindríem:

$x_1 = 5$	$y_1 = 3$	$n_{11} = 2$	
$x_1 = 5$	$y_3 = 7$	$n_{13} = 1$	
$x_2 = 6$	$y_2 = 5$	$n_{22} = 2$	La resta de $n_{i,j} = 0$

Per una altra part, és evident que tenir tres-cents parells ordenats d'observacions aclareix ben poc la informació. No és possible observar gairebé res. En conseqüència, és necessari representar les dades de manera que hi siguin més comprensibles i en faciliten l'estudi. La manera de fer-ho és mitjnaçant taules (**taules 1 i 2**).

Taula 1

X	Y	nij
x1	y1	n11
...	…	…
xi	yj	nij
.	.	.
. xh	yk	nhk

Per a construir aquesta taula ordenem una de les variables, per exemple X, i anem associant-li el valor corresponent de la variable Y, així com la seua freqüència absoluta conjunta. Si les dades foren agrupades en intervals, aleshores la representació mitjançant aquesta taula es faria de forma similar. En ocasions s'utilitzarà la marca de la classe com a representació de l'interval.

Aquesta taula presenta i ordena les dades, no obstant, en algunes ocasions no és la taula més adequada i cal construir la taula de doble entrada o de contingència.

Exemple 1

Valor del terreny	7	7	7	6,9	6,9	5,5	3,7	3,7	5,9	3,8	3,8	3,8	8,9	8,9	9,6
Cost de la vivenda	67	67	67	63	63	60	54	54	58	36	36	36	76	76	87

Valor del terreny	9,9	9,6	9,9	10	10	5,9	3,8	9,6	9,6	8,9	3,7	5,5	3,8	8,9	9,9
Cost de la vivenda	89	87	89	92	92	58	36	87	87	76	54	60	36	76	89

L'any 1999 els residents d'un petit poble estaven preocupats per l'increment del cost de la vivenda en la zona. L'alcalde considerava que els preus de la vivenda fluctuaven amb els preus dels solars. Els costos dels terrenys i els de les vivendes (en milers d'euros) sobre els quals es van construir les cases són els següents:

X	Y	$n_{i,j}$
3,7	54	3
3,8	36	4
5,5	60	2
5,9	58	2
6,9	63	2
7	67	2
7	67,15	1
8,9	76	4
9,6	87	4
9,9	89	3
10	92	2

Com es pot apreciar, les dades recollides en la taula anterior aporten poca informació. Es va construir, ja que no hi ha molts parells diferents, la taula amb les tres columnes. Se suposarà:

X: Valor del terreny.
Y: Valor de la vivenda.

Taules de doble entrada o de contingència

La taula anterior, així com s'ha comentat abans, algunes vegades és incòmoda i és preferible utilitzar la taula de doble entrada; la qual permet extraure molta més informació de la distribució de dades. La *taula 2* presenta la forma de rectangle, tal com es pot observar tot seguit:

Taula 2

Y / X	y1	y2	.....	yj	…..	yk	ni.
x1	n11	n12		n1j	…..	n1k	n1.
x2	n21	n22	…..	n2j	….	n2k	n2.
	..	.	.		.		.
xi	ni1	ni2		nij	….	nik	ni.
Xh	nh1	nh2		nhj	..	nhk	nh.
n. j	n. 1	n. 2		n. j	…	n. k	n

En la primera fila se situen les diferents categories o valors que prenen una de les components, i en la primera columna els valors o les categories relatives a la segona variable (si és possible, ordenades tant la fila com la columna). D'aquesta forma, qualsevol nombre que apareix a una cel·la interior de la taula de doble entrada, és la freqüència absoluta conjunta de la dada bivariant formada pels valors corresponents ubicats a la primera fila i a la primera columna. En algunes ocasions també se sol representar en cada cel·la la freqüència relativa conjunta, a més a més de l'absoluta.

D'altra banda, els valors que apareixen a l'ultima columna i l'última fila corresponen a les freqüències absolutes dels valors de les variables de la primera columna i la primera fila respectivament. Així, *ni.* representa la freqüència absoluta del valor de la variable X *xi*.

Si les dades foren agrupades en intervals, aleshores la representació mitjançant aquesta taula es faria de forma similar, utilitzant la marca de la classe com a representació de l'interval.

Exemple 2

Amb les mateixes dades que a l'anterior exemple 1, la taula de doble entrada queda:

X \ Y	36	54	58	60	63	67	67,15	76	87	89	92	ni.
3,7		3										3
3,8	4											4
5,5				2								2
5,9			2									2
6,9					2							2
7						2	1					3
8,9								4				4
9,6									4			4
9,9										3		3
10											2	2
n. j	4	3	2	2	2	2	1	4	4	3	2	29

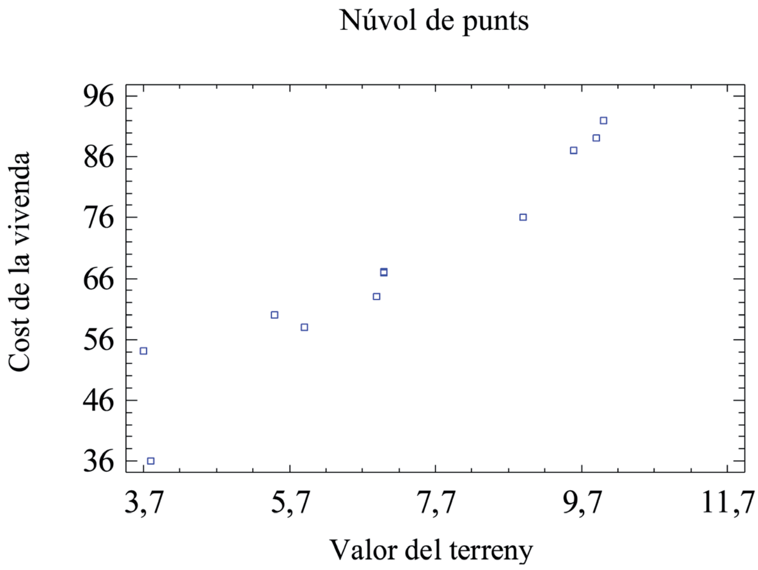
Les cel·les buides representen unes freqüències absolutes conjuntes iguals a zero. De la mateixa manera que ocorria amb les distribucions de dades unidimensionals, les representacions gràfiques faciliten la comprensió de la distribució amb tan sols un colp d'ull.

Representacions gràfiques: diagrama de dispersió o núvol de punts

La representació gràfica de la distribució de freqüències d'una variable bidimensional (X,Y) varia sensiblement segons la naturalesa de les variables. Si les variables són discretes, la representació comú de la distribució conjunta és el núvol de punts o diagrama de dispersió, el qual es construeix situant sobre l'eix horitzontal d'un sistema cartesià els diferents valors de la variable X, sobre la vertical els de la variable Y, i un punt en la posició (x_i, y_j) si és que aquesta observació té una freqüència absoluta conjunta d'1. Si en tinguera més d'1, hi hauria diferents possibilitats per a representar-ho: dibuixar punts de diferents superfícies (la qual representarà la freqüència), escriure la freqüència al costat del punt marcat, etc.

Exemple 3

Amb les dades de l'exemple 1 que s'està considerant relatiu al valor del terreny i el cost de la vivenda, el núvol de punts és:



Distribucions estadístiques marginals i condicionades

És evident que la taula de doble entrada esmentada a l'epígraf anterior ofereix molta informació. De fet, és possible analitzar cada variable component de la variable conjunta, així com una variable condicionada a un valor concret de l'altra.

Distribucions marginals

Si les variables X i Y són no agrupades o qualitatives, la distribució marginal de X s'obté de la taula de doble entrada adjuntant a cada un dels valors $x_1, x_2, \ldots\ldots, x_h$ de la variable estadística X, les seues freqüències absolutes, que hi vénen donades a l'última columna de la taula. Així mateix, s'obté la distribució marginal de Y. En aquest cas els valors de la variable $y_1, y_2, \ldots\ldots, y_k$ i les seues freqüències absolutes apareixen en la primera i l'última fila respectivament. Si les variables foren agrupades en intervals, es faria el mateix procediment prenent la marca de la classe com a representant de l'interval, i per tant, com a valor de la variable estadística. Cal dir que cada distribució marginal pot ser tractada estadísticament com una variable unidimensional.

Exemple 4

En l'exemple 1 que s'està considerant del valor del terreny i el cost de la vivenda:

Distribució marginal: Cost de la vivenda	
Y	n. j
36	4
54	3
58	2
60	2
63	2
67	2
67,15	1
76	4
87	4
89	3
92	2
	29

Distribució marginal: Valor del terreny	
X	ni.
3,7	3
3,8	4
5,5	2
5,9	2
6,9	2
7	3
8,9	4
9,6	4
9,9	3
10	2
	29

Distribucions condicionades

De la taula de doble entrada, també és possible obtenir, a més de les distribucions marginals, unes altres distribucions. Si s'associa als valors de Y les freqüències corresponents a la fila en què està ubicat el valor $\mathbf{x_i}$ de X, resulta la distribució condicionada de Y a $\mathbf{x_i}$ (Distribució de la variable $Y/_{X = x_i}$). Anàlogament, però tenint present les columnes en lloc de les files, s'obtindria la distribució de X condicionada a $\mathbf{y_j}$ de Y (Distribució de la variable $X/_{Y = y_j}$).

Exemple 5

Si en l'exemple 1 es desitja conèixer la distribució del preu la vivenda quan el preu del solar és de 7000 euros, la distribució condicionada és $Preu\ vivenda/_{Preu\ solar = 7000}$ es pot obtenir de la taula de doble entrada:

Distribució condicionada: Preu solar 7.000 €	
$Y/X = 7$	$n_{.j}$
67	2
67,15	1
	3

Correlació lineal

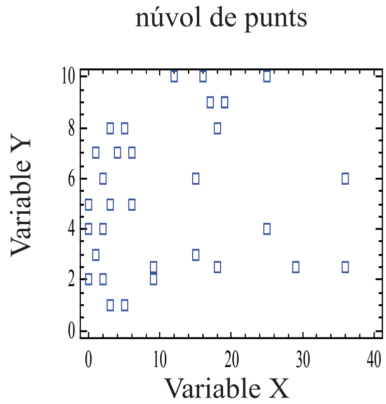
Quan s'estudien dues variables estadístiques conjuntament, és important saber si hi ha algun tipus de relació entre elles. Així, si es recolliren tres-centes dades en què la primera variable fóra l'altura d'una persona i la segona, el resultat de llançar un dau, segurament la intuïció diria que les dues variables no tenen cap tipus de relació entre si. Si per contra, es consideraren les variables hores extra que treballa una persona i sou que cobra mensualment, la relació canviaria, fins al punt de conèixer el sou d'un individu si se sap les hores extres que hi fa. Es podria dir que les dues variables estan lligades per una relació funcional. Tanmateix, si es consideren les variables hores de preparació d'un examen i nota obtinguda, la intuïció establiria que sí hi ha alguna relació entre totes dues variables, sent molt més forta que el primer cas, però més dèbil que en el segon.

Com és evident, les relacions funcionals gaudeixen d'una fórmula que demostra el tipus de relació. Pel contrari, per a la resta de parells de variables no hi ha cap fórmula absoluta, malgrat els lligams que existeixen en alguns casos. És per evidenciar-ho que sorgeix el concepte de correlació, r_{XY}. Així, si dues variables hi tenen una relació molt forta el valor absolut de la correlació serà molt pròxim a 1. En cas contrari serà pròxim al zero. Els casos 0 i 1, equivalen a no tenir cap tipus de relació i a comptar amb una relació funcional. El vector següent ho resumeix:

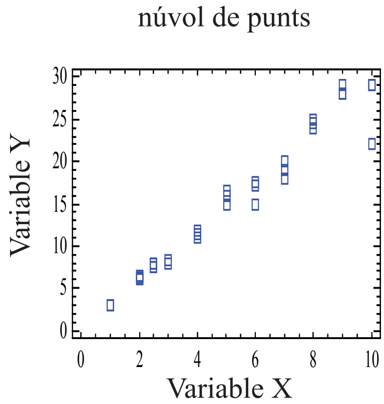
$\lvert r_{XY} \rvert = 0$	$0 < \lvert r_{XY} \rvert < 1$	$\lvert r_{XY} \rvert = 1$
No existeix relació	Existix relació	Relació funcional

Quan el tipus de relació funcional que s'estudia entre les variables és una funció lineal (una funció del tipus $y = ax + b$), es parla de correlació lineal. Al llarg de la unitat, quan es mencione el terme correlació es considerarà la correlació lineal, si no s'explicita una altra cosa.

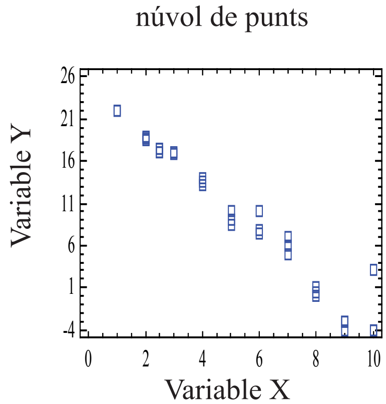
Una primera manera d'observar la relació existent entre les variables X i Y és mitjançant els gràfics de dispersió. Així, tenint en compte l'exposat al començament d'aquest punt sobre la correlació:



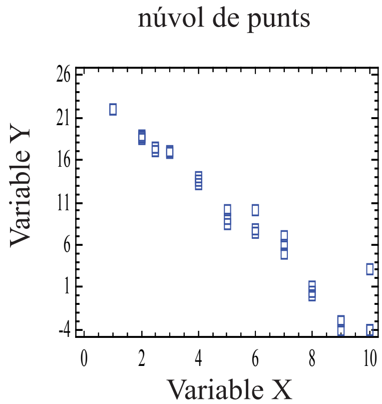
a) No existix correlació



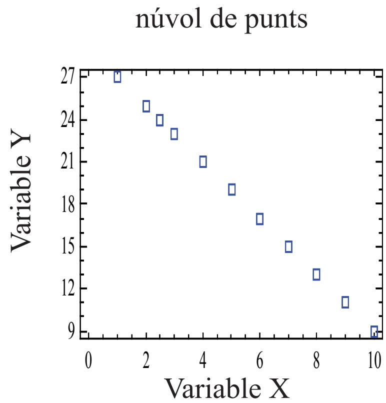
b) Correlació lineal positiva marcada



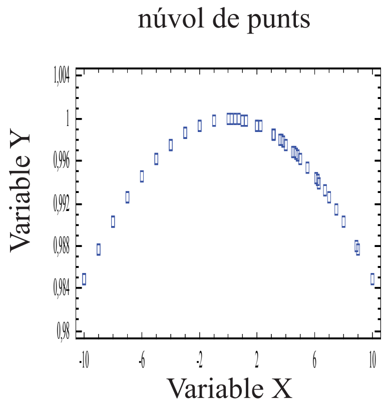
c) Correlació lineal positiva perfecta



d) Correlació lineal negativa marcada



e) Correlació lineal negativa perfecta



f) Correlació no lineal entre X i Y

Com es pot observar, a l'exemple *e*) es detecta una relació entre les variables X i Y. No obstant, és evident que no es tracta d'una relació lineal; ja que aquests tipus de relacions determinen un núvol de punts semblants a una línia recta (exemples *b*, *c*, *d* i *e*). A l'exemple *a*), no es distingeix cap tipus de lligam entre totes dues variables; els punts estan molt dispersos.

Covariància

El gràfic és una primera aproximació a l'estudi de la relació que existeix entre les variables, però únicament aporta informació de caire intuïtiu. El concepte que és necessari definir per a poder decidir si hi ha o no relació lineal entre dues variables és el de *correlació lineal*. En primer lloc però, cal introduir el concepte de covariància.

La covariància és un estadístic (o un paràmetre) que es calcula semblantment al de variància i permet conéixer si dues variables estan relacionades o no linealment, es representa per S_{XY} i es calcula segons la fórmula:

$$S_{XY}=\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{(x_i-\overline{X})(y_j-\overline{Y})\cdot n_{ij}}{n}$$

La interpretació d'aquest estadístic és la següent:

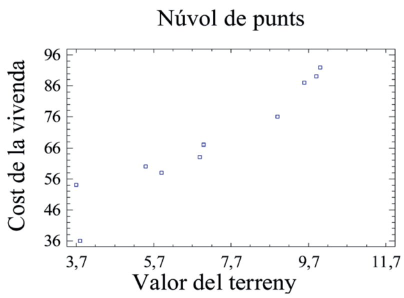
- Si $S_{XY} > 0$ ==> dependència lineal directa (positiva), és a dir, a grans valors de X corresponen grans valors de Y (Exemples *b*) i *c*)).
- Si $S_{XY} = 0$ ==> incorrelacionades, és a dir, no hi ha relació lineal (Exemples *a*) i *f*)).
- Si $S_{XY} < 0$ ==> dependència lineal inversa o negativa, és a dir, a grans valors de X corresponen menuts valors de Y (Exemples *d*) i *e*)).

Exemple 6

En el cas que s'està considerant del valor del terreny i el valor de la vivenda, la covariància és:

Sabent que Valor terreny: $=\overline{X} = 7{,}1586$, Valor vivenda: $=\overline{Y} = 68{,}0052$ i n = 29 es calcula:

$$S_{XY}=\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{(x_i-\overline{X})(y_j-\overline{Y})\cdot n_{ij}}{n}=\sum_{i=1}^{10}\sum_{j=1}^{11}\frac{(x_i-7{,}1586)(y_j-68{,}0052)\cdot n_{ij}}{29}=\underline{42{,}1527\ €^2}$$



Així doncs, el valor de la covariància és coherent amb el núvol de punts que s'ha obtingut: sembla existir un relació directa o positiva entre valor del terreny i cost de la vivenda. A més valor del terreny, més cost de la vivenda.

Propietats de la covariància:

- Si a tots els valors de la variable X, se'ls hi suma una constant *b* i a tots els valors de la variable Y una constant C, la covariància no varia.

  És a dir, $S_{X+b\ Y+C} = S_{XY}$

- Si tots els valors d'una variable X es multipliquen per una constant *a* i tots els valors de la variable Y per una constant *b*, la covariància queda multiplicada pel producte de les constants.

  És a dir, $S_{a \cdot X\, b \cdot Y} = a \cdot b\, S_{XY}$

- A partir de les anteriors: si es tenen dues variables X i Y amb covariància S_{XY}, i dues transformacions lineals de les variables de la forma X' = *ax* + *c*, i Y' = *by* + *d*, la nova covariància es relaciona amb l'anterior de la forma:

  És a dir, $S_{a \cdot X\, b \cdot Y} = a \cdot b\, S_{XY}$

Càlcul de la covariància

Existeix una altra forma de d'obtenir la covariància, la qual és de càlcul més senzill:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y}$$

Es pot demostrar l'equivalència d'ambdues definicions mitjançant procediments algebraics elementals.

Exemple 7

Es veurà tot seguit un exemple d'aplicació d'aquesta darrera propietat. De la següent taula de doble entrada es determinarà la covariància:

X \ Y	1,6	1,7	1,8	
60	2	1	0	3
70	2	4	2	8
80	1	1	4	6
90	0	2	1	3
	5	8	7	20

En primer lloc es calcularà la mitjana aritmètica de cada variable marginal:

$$\overline{X} = 74{,}5 \text{ i } \overline{Y} = 1{,}71$$

En segon lloc, cal dterminar el primer sumand de S_{XY}, $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per això, és necessari calcular primer els productes, sumar-los tots i després dividir el resultat pel nombre total de dades:

60 * 1,6 * 2 =	192	Per tant es tindrà:
60 * 1,7 * 1 =	102	
70 * 1,6 * 2 =	224	$\sum_{i=1}^{h}\sum_{j=1}^{k} x_i \cdot y_j \cdot n_{ij} = 2554$
70 * 1,7 * 4 =	476	
70 * 1,8 * 2 =	252	$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{2554}{20}$
80 * 1,6 * 1 =	128	
80 * 1,7 * 1 =	136	
80 * 1,8 * 4 =	576	$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} =$
90 * 1,7 * 2 =	306	
90 * 1,8 * 1 =	162	$= \frac{2554}{20} - 74{,}5 \cdot 1{,}71 =$
TOTAL SUMA =	2554	= 127,7-127,395 = 0,305

Correlació lineal

La covariància permet discernir si dues variables X i Y tenen una relació positiva, negativa o zero, però no aporta informació del grau de dependència d'una variable respecte a l'altra (referències bibliogràfiques 6, 10 i 17). A més a més, la covariància depén de les unitats de mesura emprades per a X i Y (si per exemple X es mesura en m^3 i Y en mm^3, cada desviació de X augmenta S_{XY} 10^9 vegades). Per a fer front a aquestes dues dificultats es definix el concepte ja introduït anteriorment de correlació lineal r_{XY}:

$$r_{XY} = \frac{S_X}{S_X \cdot S_Y}$$ sent S_X i S_Y les desviacions típiques de X i Y.

És evident que, per definició, el coeficient de correlació lineal informa de les mateixes coses que la covariància. A més a més, compleix una propietat molt important, està fitat per 1 i per –1. Així doncs, r_{XY} es caracteriza per:

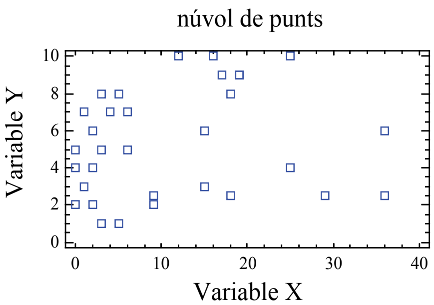
- Ser adimensional i estar sempre comprés entre –1 i 1.
- Si hi ha relació lineal forta positiva, $r_{XY} > 0$ i està a prop de 1.
- Si hi ha relació lineal negativa forta, $r_{XY} < 0$ i està a prop de –1.
- Si no hi ha relació lineal r_{XY} serà 0.

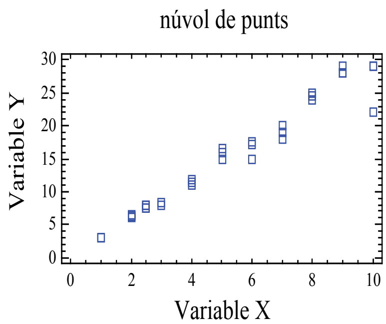
Exemple 8

En l'exemple número 1 que s'està considerant, per a calcular la correlació és necessari primer conèixer les variàncies i la covariància. Aprofitant els càlculs anteriors, es té: $S_{XY} = 42{,}1527$; $S_X = 18{,}1656$; $S_Y = 2{,}4242$ En conseqüència, el coeficient de correlació lineal $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{42{,}1527}{18{,}1656 \cdot 2{,}4242} = 0{,}9572$. Per tant, la relació lineal entre les dues variables és alta.

Recta de regressió. Qualitat de l'ajust

Com s'ha exposat anteriorment, quan s'estudien dos característiques simultàniament sobre una mostra, es pot considerar que una d'elles influeix sobre l'altra d'alguna manera. L'objectiu principal de la regressió és descobrir el mode en què s'hi relacionen. Un dibuix del núvol de punts o diagrama de dispersió de la distribució pot indicar si és raonable pensar que pot haver-hi una bona correlació lineal entre les dues variables.





En els diagrames anteriors es pot observar com en el de la dreta, una línia recta pot aproximar-se a quasi tots els punts, mentre que en l'altre, qualsevol recta deixa molts punts allunyats d'ella. Així, fer una anàlisi de regressió lineal només estaria justificat en l'exemple de l'esquerra.

Com es pot veure en ambdós diagrames, cap recta és capaç de passar per tots els punts. De totes les rectes possibles, la recta de regressió de Y Sobre X és aquella que minimitza «l'error d'aproximació», considerant X com a variable explicativa o independent i Y com l'explicada o dependent. Però, com es calcula la recta i es minimitza l'error?

núvol de punts

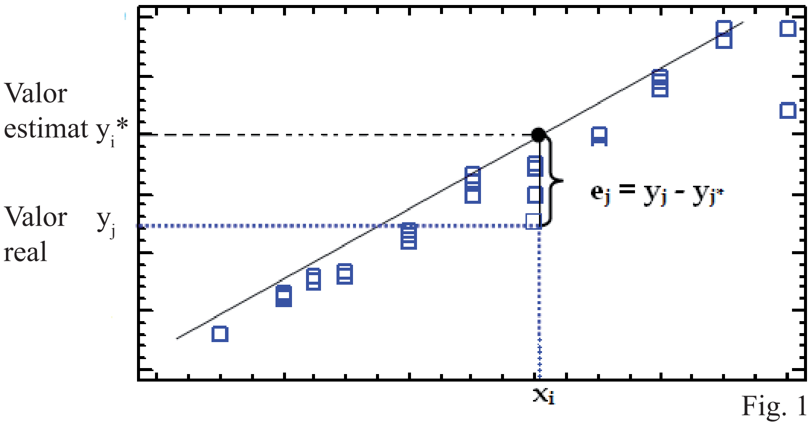


Fig. 1

Es considera la recta $y = a + b$ X on a i b son paràmetres. D'aquesta manera, la recta o funció lineal és genèrica (representa a totes les funcions lineals possibles, únicament cal donar valors als paràmetres per obtenir les infinites rectes). Allò que es realitzarà consisteix a trobar els valors dels paràmetres a i b de manera que la recta s'ajuste tant com siga possible als punts de la Fig. 1. El mètode que s'empra per a cercar els valors dels paràmetres a i b és el dels mínims quadrats.

Usant tècniques de derivació es dedueix que, de tots els possibles valors de a i de b, aquells que minimitzen la suma anterior són:

$$a = \bar{y} - \frac{s_{xy}}{s_x^2} \cdot \bar{x} \qquad \text{i} \qquad b = \frac{s_{xy}}{s_x^2}$$

(NOTA: Cal no oblidar, que si es coneixen les dades també es coneixen els termes: $\left\{\bar{y}, \bar{x}, S_{xy}, S_x^2\right\}$, i per tant a i b seran nombres reals en el moment que es produisquen les substitucions.)

Així, substituint Y = $a + b$ X, l'equació de la recta de regressió de Y sobre X és:

$$y = \left(\bar{y} - \frac{s_{xy}}{s_x^2} \cdot \bar{x}\right) + \left(\frac{s_{xy}}{s_x^2}\right) \cdot x$$

i recol·locant els termes es pot escriure de la forma:

$$y - \bar{Y} = \frac{S_{XY}}{S_X^2} \cdot (x - \bar{X})$$

Si s'haguera pres Y com a variable independent o explicativa, i X com a dependent o explicada, la recta de regressió que es necessitaria és la que minimitza errors de la X. S'anomena recta de regressió de X sobre Y i es calcula fàcilment permutant els llocs de x i y, obtenint-se:

$$x - \overline{x} = \frac{s_{xy}}{s_y^2} \cdot \left(y - \overline{y}\right)$$

Exemple 9

En l'exemple que s'està considerant, la variable independent és el valor del terreny i el valor de la vivenda, la variable dependent.

Pels estudis realitzats al llarg de la unitat se sap que la relació és directa, doncs la covariància és positiva. Com que la correlació obtinguda ha estat un nombre proper a 1, la relació lineal entre les dues variables és marcada. Per tant, la recta en el càlcul de la recta de regressió té sentit. Per a calcular-la, s'utilitzarà la darrera expressió:

$y - \overline{Y} = \frac{S_{XY}}{S_X^2} \cdot (x - \overline{X})$. Així, $\frac{S_{XY}}{S_X^2} = \frac{42{,}1527}{5{,}8768}$ i la recta serà:

$y - 7{,}1586 = \frac{42{,}1527}{5{,}8768}$ $(x - 68{,}0052)$, aïllant y, la recta $y = 16{,}6583 + 7{,}17274\, x$.

Per tant:

Cost de la vivenda = 16,6583 + 7,17274 · valor del terreny

Qualitat de l'ajust. Coeficient de determinació

El coeficient de determinació lineal es pot definir com el percentatge de variància de Y que es pot explicar per X i, se'l sol anomenar *Qualitat o bondat de l'ajustament* perquè valora la proximitat del núvol de punts a la recta de regressió (o dit d'una altra manera, com està d'ajustat el núvol de punts a la recta de regressió).
Pel que fa al càlcul del coeficient de determinació, cal definir prèviament:

- La variància de la variable Y que és explicada per la regressió lineal, anomenada S_r^2, i que representa la variabilitat de la variable Y causada per les variacions de la variable X.

- La variància residual, que es representa per S_e^2, determina en quina mesura difereixen els valors ajustats per la recta dels valors observats. És a dir, es planteja mesurar la magnitud dels residus.

Així:

$$S_r^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}(y_j^* - \overline{y^*})^2 \frac{n_{ij}}{n} \qquad i \qquad S_e^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}(y_j - y_j^*)^2 \frac{n_{ij}}{n}$$

Es pot demostrar matemàticament que, en la regressió lineal de la variable Y sobre la variable X, la variància de la variable Y es pot descompondre de la manera següent:

$$S_Y^2 = S_r^2 + S_e^2$$

Així doncs, de la relació es dedueix que com major siga la variància explicada per la regressió lineal (S_r^2) respecte de la variància total, menor serà la variabilitat de l'error d'ajustament (S_e^2) i millor serà la bondat de l'ajustament.

Si ara es divideix l'expressió anterior per S_Y^2, s'obté: $1 = \frac{S_r^2}{S_Y^2} + \frac{S_e^2}{S_Y^2}$

I reprenent el significat del *coeficient de correlació lineal (R^2)* com el percentatge de variància de Y que es pot explicar per X, es té :

$$R^2 = \frac{S_r^2}{S_Y^2} = 1 - \frac{S_e^2}{S_Y^2} \qquad \text{(En tant per un)}$$

D'aquesta definició es poden treure algunes conclusions:

- $0 \leq R^2 \leq 1$, per ser la part d'un total.
- $R^2 = 1$ implica que la variància residual és nul·la i per tant l'ajust és perfecte. En conseqüència, la relació entre totes dues variables és lineal.
- $R^2 = 0$ implica que la variància residual és igual a la variància de la variable Y i que la variable explicativa no aporta informació vàlida per a l'estimació de la variable explicada. En conseqüència, no existeix relació lineal entre les dues variables.
- Com més pròxim a 1 estiga R^2 millor serà la bondat o qualitat de l'ajustament.

Per una altra part, en una regressió lineal es pot demostrar que $R^2 = \frac{S_R^2}{S_y^2} = \frac{S_{XY}^2}{S_x^2 S_y^2}$

Que evidentment coincideix amb el quadrat del coeficient de correlació lineal i en justifica totes les propietats abans esmentades d'ambdós coeficients.

$$r_{XY}^2 = R^2$$

Prediccions. Usos i abusos

El primer objectiu de la regressió lineal era posar de manifest la relació existent entre dues variables estadístiques. Una vegada es constata que n'hi ha, i es calcula la recta de regressió apropiada, aquesta es pot usar per a obtenir valors de la variable explicada, a partir de valors de la variable explicativa.

Per exemple, si es comprova una bona correlació lineal entre les variables X = «hores d'estudi setmanal» i Y = «nota de l'examen», amb una recta de regressió (de Y sobre X) igual a $Y = 0{,}9 + 0{,}6x$ es pot plantejar la pregunta següent:

Quina nota pot obtenir (segons les dades) un alumne que estudia 10 hores setmanals?

I la resposta és tan senzilla com calcular Y, substituint a l'equació de la recta $x = 10$, resultant $Y = 6{,}9$. El coeficient determinació és la dada que indicarà si la predicció obtinguda és *fiable* o no, ja que és el coeficient que informa sobre la qualitat de l'ajust.

En el moment de fer prediccions cal tenir certes precaucions, perquè és possible que obtenir resultats absurds. Segons la recta de regressió anterior, un alumne que estudie 20 hores per setmana ($x = 20$) tindria un resultat de 12,9 punts en el seu examen, la qual cosa no té sentit si s'avalua sobre 10. La limitació de la predicció consisteix en el fet que *només es pot realitzar per a valors de X que estiguen situats dintre del rang dels valors de X.*

Objectius

Els problemes han de permetre que l'alumnat assoleixen els objectius didàctics següents:

2*a*) Saber analitzar i extraure informació d'una distribució de dades bidimensional a partir de la construcció de la taula de doble entrada.
2*b*) Saber extraure conclusions de l'anàlisi tant de les distribucions marginals com de les condicionades d'una distribució de dades bidimensional.
2*c*) Distingir gràficament i analíticament si les dues variables d'una distribució de dades bidimensional tenen relació lineal.
2*d*) Saber calcular i interpretar la covariància, així com aplicar les propietats que aquest estadístic compleix.
2*e*) Saber calcular el coeficient de correlació lineal així com la seua interpretació.
2*f*) Construir la recta de regressió lineal d'una variable estadística respecte de l'altra en una distribució de dades bidimensional.
2*g*) En una distribució de dades bidimensional saber predir el valor d'una variable a partir d'un valor de l'altra mitjançant la recta de regressió i conéixer-ne la fiabilitat.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

OBJECTIUS \ EXERCICIS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2a	X	X	X	X															
2b	X	X	X	X															
2c					X	X	X												
2d						X	X												
2e								X	X	X	X	X	X	X	X	X	X	X	X
2f											X	X	X	X	X	X	X	X	X
2g											X	X	X	X	X	X	X	X	X

Enunciats

2*a*) Saber analitzar i extraure informació d'una distribució de dades bidimensional a partir de la construcció de la taula de doble entrada.
2*b*) Saber extraure conclusions de l'anàlisi tant de les distribucions marginals com de les condicionades d'una distribució de dades bidimensional.

Exercici 1

Una empresa ha entrevistat vint-i-cinc dels seus treballadors amb tasques administratives per a conèixer el grau d'implicació en la seua formació professional. A cadascun li preguntà el nombre de cursos de formació de més de 30 hores i el nombre de cursos de perfeccionament d'idiomes que havia realitzat en els darrers tres anys. Els resultats són els que es mostren en la taula següent:

formació	8	9	4	5	6	7	7	9	10	7	5	6	7
idiomes	8	8	3	5	7	7	8	10	10	7	6	7	8

formació	8	5	8	9	8	8	7	7	9	9	8	7
idiomes	7	5	8	8	7	8	7	7	8	10	8	8

a) Construïu la taula de freqüències conjunta.
b) Calculeu el nombre mitjà de cursos de formació i el nombre mitjà de cursos d'idiomes que han realitzat els treballadors de l'empresa.
c) Calculeu el nombre mitjà de cursos de formació que han efectuat aquells treballadors que n'han fet set de perfeccionament dels idiomes.
d) Quina proporció de treballadors han realitzat més de cinc cursos en ambdues categories? Quina proporció de treballadors ha fet més de cinc cursos de formació? I més de cinc cursos d'idiomes?
e) Quina proporció de treballadors han realitzat més de set cursos de formació i més de vuit en idiomes?
f) Quin percentatge dels treballadors que han fet cinc cursos o més de formació, han fet set cursos o més d'idiomes?

Exercici 2

Una empresa vol obrir un punt de venda en un barri d'una gran ciutat de la Comunitat Valenciana. Com que el segment de població a qui va adreçat el producte és a persones d'edats compreses entre 45 i 55 anys ha decidit enquestar una mostra de 50 veïns en aquesta franja del barri. La taula següent exposa dues de les preguntes que apareixien en l'enquesta: edat i ingressos mensuals en milers d'euros.

Edat	50	51	53	50	51	48	50	49	52	52	49	50	52	51	52	49	50
Ingressos mensuals	3.2	4.1	4.5	3	3.6	2.9	3.8	3.8	3.6	3.9	3	3.8	4.1	3.5	4.0	3.1	3.1

Edat	51	50	51	52	53	52	52	51	50	51	54	50	51	51	51	52	51
Ingressos mensuals	4.3	3.3	3.9	3.7	4.1	4.2	3.5	3.8	3.6	3.4	4.6	3.5	3.6	3.1	4	3.8	4.2

Edat	52	51	50	51	49	51	48	50	52	53	52	50	52	51	51	51
Ingressos mensuals	4	4.4	3.9	3.7	3.4	3.3	2.7	3.4	3.6	4.4	4.3	3.3	4.2	4.2	3.3	3.7

a) Construïu la taula de doble entrada agrupant els ingressos mensuals en intervals d'amplària 0,5 i de manera que l'extrem petit de la primera classe siga 2,5.

b) Quins són els ingressos mitjans dels enquestats que tenen 51 anys? Quin percentatge d'aquests té uns ingressos inferiors a 4000 euros?

c) Quina és la mitjana d'edat dels enquestats que tenen uns ingressos entre 3500 i 4000 euros? Quin percentatge d'aquests té 50 o 51 anys?

d) Quin percentatge del clients ingressa mensualment 4000 euros o més i té més de 50 anys?

e) Quin percentatge de les persones enquestades té més de 51 anys o uns ingressos de 4000 euros o més?

Exercici 3

El departament de recursos humans d'una empresa ha decidit realitzar dos tests per tal de seleccionar les persones que hauran de fer-se càrrec d'un projecte d'innovació. Les notes obtingudes pels aspirants es mostren en la taula següent:

TEST 1	7	6	5	4	5	8	7	8	9	6	5	8	6	8	7	8	7	6	6	9
TEST 2	8	7	6	6	7	10	9	9	10	8	6	10	8	9	8	8	7	8	6	8

a) Construïu la taula de doble entrada.
b) Calculeu la nota mitjana en el test 2 dels aspirants que han obtingut un 6 en el test 1.
c) Percentatge d'aspirants que obtenen un nota inferior a 8 en el test 2 d'entre aquells que obtenen un nota en el test 1 superior a 6.

Exercici 4

La taula següent mostra el nombre de persones ocupades distribuïdes atenent el sou net de l'activitat principal que desenvolupen (en centenes d'euro) i la seua edat durant l'any 2010, segons dades recollides del Ministeri de Treball i Immigració.

	SOU						
EDAT	[0, 6)	[6,10)	***[10,12)***	***[12,16)***	***[16,21)***	***[21,30)***	***[30, 40)***[1]
[16,25)	289,79	490,44	*249,08*	126,47	38,03	*1,70*	0
[25,30)	232,55	673,68	*571,85*	430,16	192,86	*20,80*	11,01
[30,45)	566,18	1777,07	*1671,91*	2190,02	1248,87	*736,77*	155,06
[45,55)	323,65	797,11	*881,81*	1123,69	724,93	*448,78*	138,99
[55,65[2]***)***	185,20	430,59	*503,77*	568,69	306,20	*225,13*	123,53

a) Construïu les taules de freqüència de les distribucions de les variables marginals i calculeu la mitjana aritmètica de cadascuna.
b) Construïu la taula de freqüència de l'edat d'aquelles persones ocupades que tenen un sou de 1200 a 1600 euros. Calculeu també l'edat mitjana de les persones que cobren entre 1200 i 1600 euros.
c) Construïu la taula de freqüència del sou d'aquelles persones ocupades que tenen 30 anys o més. Quin sou mitjà cobren?
d) Quin percentatge de persones ocupades tenen 45 anys o més i cobren 1.600 euros o més?
e) Quin percentatge de persones ocupades tenen 45 anys o més o cobren 1600 euros o més?
f) Quin percentatge d'ocupats tenen menys de 30 anys d'aquells que cobren 1200 euros o més?

1.Distingir gràficament i analíticament si les dues variables d'una distribució de dades bidimensional tenen relació lineal.
2. En la taula original el darrer interval és de 55 anys o més. S'ha tancat l'interval per a fer l'exercici.

2*c*) Distingir gràficament i analíticament si les dues variables d'una distribució de dades bidimensional tenen relació lineal.

Exercici 5

La taula següent mostra la població en edat de treballar analfabeta en la Comunitat Valenciana, Madrid, Andalusia i el País Basc al llarg dels anys 2000-2010 en milers de persones.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Andalusia	332,28	280,99	342,14	307,41	294,91	247,90	262,25	278,03	294,34	290,37	279,91
Madrid	73,15	58,88	74,06	83,33	71,83	47,35	43,35	40,34	46,68	56,98	72,21
País Basc	17,69	13,77	14,67	12,56	13,27	9,37	11,34	12,50	13,66	14,23	11,10
C. Valenciana	128,01	96,83	110,81	117,30	114,73	69,46	79,91	79,01	92,45	99,06	85,76

Font: INE

a) Representa el núvol de punts entre les variables: Població en edat de treballar a Andalusia i Població en edat de treballar a la Comunitat Valenciana. Què observeu pel que fa a la existència o no de la relació lineal entre les dues variables?
b) Representa el núvol de punts entre les variables: Població en edat de treballar a Madrid i Població en edat de treballar al País Basc. Què observeu pel que fa a l'existència o no de la relació lineal entre les dues variables?
c) Calculeu l'estadístic adequat per a confirmar les suposicions que heu fet en els dos apartats anteriors.

2*c*) Distingir gràficament i analíticament si les dues variables d'una distribució de dades bidimensional tenen relació lineal.
2*d*) Saber calcular i interpretar la covariància, així com aplicar les propietats que aquest estadístic compleix.

Exercici 6

Es van recol·lectar els valors mensuals de les despeses en publicitat d'una companyia ferroviària i el nombre de passatgers al llarg de 15 mesos. Les dades les mostra la taula:

Publicitat (en milers)	10	12	8	17	10	15	10	14	19	10	11	13	16	10	12
Passatgers (en milers)	15	17	13	23	16	21	14	20	24	17	16	18	23	15	16

a) Calculeu la despesa mitjana i el nombre mitjà de passatgers.
b) Feu-ne el núvol de punts i calculeu la covariància. És coherent el valor de l'estadístic amb el núvol de punts?
c) Si per als 15 mesos posteriors es preveu que la inversió en publicitat de cada mes augmente un 10% respecte al mateix mes del període anterior, i també es preveu que aquest fet provocarà un augment del 8% en el nombre de passatgers cada mes, quina serà la covariància en aquest segon període?

Exercici 7

Una empresa ha realitzat dos tests psicotècnics als 9 treballadors d'un departament com a part del procés de selecció del nou director del departament. La taula següent mostra els resultats obtinguts pels aspirants:

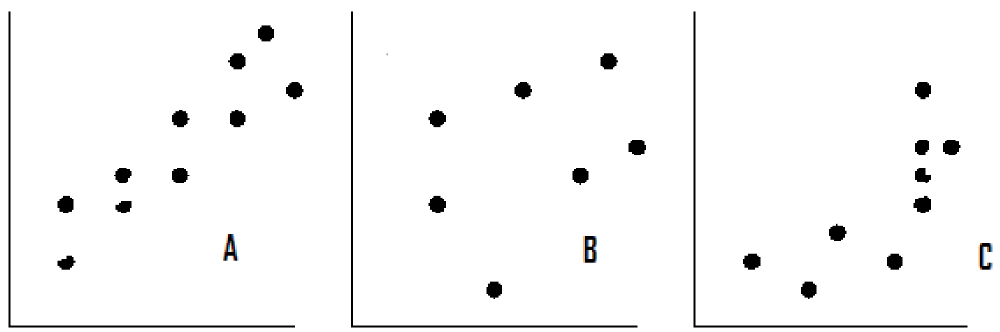
TEST 1	5	7	6	9	3	1	2	4	6
TEST 2	6	5	8	6	4	2	1	3	7

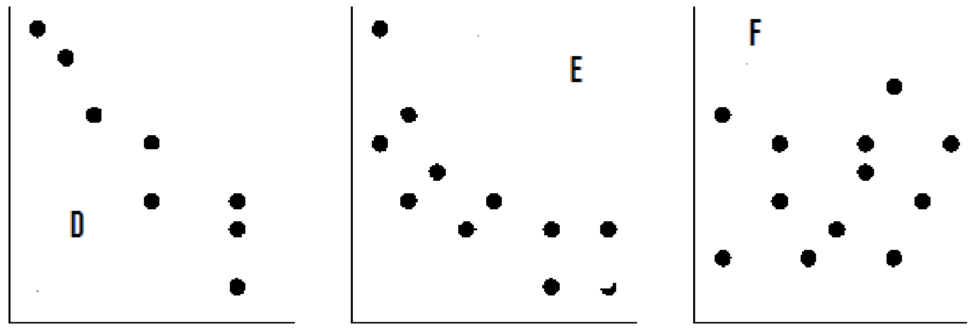
a) Calculeu la covariància. Existeix cap tipus de relació lineal entre les dues variables?
b) Hi ha hagut un error en una pregunta de cada test i el tribunal decideix augmentar un 5% la puntuació de cada participant. Calculeu-ne novament la covariància.

2*e*) Saber calcular el coeficient de correlació lineal així com la seua interpretació.

Exercici 8

Donats els següents núvols de punts, contesteu:





a) Associeu cada núvol de punts amb el valor del coeficient de correlació que li correspon: –0,9; 0,4; 0,95; –0,65; 0,1; 0,6. Raoneu la resposta.
b) Indiqueu per a cada núvol de punts el signe de la covariància i digueu quin n'és el significat.

Exercici 9

La taula següent mostra la despesa total mitjana, la despesa mitjana en aliments i begudes no alcohòliques i la despesa en habitatge, aigua, electricitat, gas i altres combustibles en euros, per nombre de persones que formen la unitat familiar l'any 2009[3] segons dades de l'INE.

Nombre de membres de la llar	*1*	*2*	*3*	*4*	*5*	*6 o més*
Despesa mitjana total	18355,25	27755,08	33414,09	38576,14	40699,09	41562,31
Despesa en habitatge, aigua, electricitat, gas i altres combustibles	7493,88	8990,72	9205,13	9645,19	10114,49	9272,18

a) Existeix una forta relació lineal entre el nombre de membres que viuen en una llar i la despesa mitjana total? Raoneu la resposta.
b) I entre el nombre de membres que viuen en una llar i la despesa en habitatge, aigua, electricitat, gas i altres combustibles? Raoneu la resposta.

3. Per a fer l'exercici, considera 7 membres a l' interval 6 o més membres de la llar.

Exercici 10

El director de recursos humans d'una empresa ha realitzat dos tests psicotècnics per a seleccionar les persones que han de treballar en el departament de màrqueting. S'hi han presentat 9 persones i els resultats obtinguts en cadascun dels tests han estat els següents:

TEST 1	175	181	192	211	235	255	275	286	292
TEST 2	169	185	202	219	240	266	295	329	357

Tenint en compte els resultats dels tests, creieu que el director podria haver eliminat un dels dos tests per a decidir els candidats? Raoneu la resposta.

2*f*) Construir la recta de regressió lineal d'una variable estadística respecte de l'altra en una distribució de dades bidimensional.
2*e*) Saber calcular el coeficient de correlació lineal així com la seua interpretació.
2*g*) En una distribució de dades bidimensional saber predir el valor d'una variable a partir d'un valor de l'altra mitjançant la recta de regressió i conèixer-ne la fiabilitat.

Exercici 11

En una mostra de 150 empreses del sector de serveis es recullen dades sobre el nombre de treballadors de l'empresa (X) i la facturació (Y) anual en milions d'euros. Els resultats es mostren resumits en els següents estadístics:

$$\overline{X} = 14\, treballadors;\ \overline{Y} = 100\, milions;\ S_X = 2\, treballadors;\ S_Y = 25\, milions;$$
$$S_{XY} = 45\, treballadors \times milió$$

a) Calculeu la correlació lineal i interpreteu-la.
b) Calculeu el model de regressió lineal que millor aproxima la facturació en funció del nombre de treballadors.
c) En funció d'aquest ajust, calculeu de manera aproximada la quantitat que s'espera que facture una empresa amb 15 treballadors. És fiable aquesta predicció? Raoneu la resposta.
d) Calculeu el model de regressió lineal que millor aproxima el nombre de treballadors en funció de la facturació.
f) En funció d'aquest ajust, calculeu de manera aproximada el nombre de treballadors que s'espera que tinga una empresa que facture 105 milions. És fiable aquesta predicció? Raoneu la resposta.

Exercici 12

Les dues taules següents mostren el grau mitjà de satisfacció dels ocupats segons el treball que realitzen per edat i pel nivell d'estudis l'any 2010. Les dades estan extretes del Ministeri de Treball i Immigració.

NIVELL ESTUDIS	*GRAU DE SATISFACCIÓ*
1	7,05
2	7,09
3	7,21
4	7,23
5	7,50
6	7,55

EDAT	*GRAU DE SATISFACCIÓ*
[16,25)	7,33
[25,30)	7,39
[30,45)	7,37
[45,55)	7,30
[55,65)[4]	7,43

Cal dir que la variable nivell d'estudis ha estat convertida a numèrica discreta per ser graduable. Així l'equivalència és*: 1 = menys que primaris; 2 = primaris; 3 = secundaris; 4 = batxillerat; 5 = formació professional i 6 = universitaris.* Aquesta conversió s'ha fet a efectes didàctics:

a) Calculeu el coeficient de relació lineal de totes dues parells de variables. En quina de les dues convindria calcular la recta de regressió?
b) Calculeu la recta de regressió del grau de satisfacció en funció del nivell d'estudis.

Exercici 13

El grau mitjà de satisfacció mitjà dels ocupats segons el treball que realitzen per nivell d'ingressos i per sexe l'any 2010 es mostra en la taula següent. Les dades estan extretes del Ministeri de Treball i Immigració.

NIVELL D'INGRESSOS	*GRAU DE SATISFACCIÓ HOMES*	*GRAU DE SATISFACCIÓ DONES*
[0,600)	6,19	7,253
[600,1000)	6,83	7,234
[1000,1200)	7,28	7,339
[1200,1600)	7,39	7,61
[1600,2100)	7,60	7,768
[2100,3000)	7,82	7,682
[3000,4000)[5]	7,925	7,499

4. En la taula original el darrer interval és 55 anys o més. S'ha tancat l'interval per a fer l'exercici.
5. En la taula original el darrer interval és 3000 euros o més. S'ha tancat l'interval per a fer l'exercici.

a) Calculeu el coeficient de correlació lineal entre les variables Nivell d'ingressos i Grau de satisfacció en els homes i entre les variables Nivell d'ingressos i Grau de satisfacció en les dones. Quines conclusions es poden obtenir?
b) Calculeu la recta de regressió que explique el grau de satisfacció mitjà en el treball dels homes en funció del nivell d'ingressos.

Exercici 14

El nombre total d'expedients de regulació del treball al llarg dels anys 2001-2010, segons les dades extretes del Ministeri de Treball i Immigració són les que es mostren en la taula.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Alacant	139	169	224	268	292	180	164	393	939	679
Castelló	49	59	55	76	88	59	58	291	939	777

a) Existeix cap tipus de relació lineal entre les variables? És forta aquesta relació? Raoneu les respostes.
b) Calculeu la recta de regressió lineal que relaciona el nombre d'expedients totals a Castelló en funció dels d'Alacant.

Exercici 15

La taula següent mostra el nombre total d'hipoteques signades, així com la taxa d'atur a l'Estat espanyol en el període 2004-2010, segons les dades extretes de l'INE.

	Hipoteques	***Taxa d'atur***
2004	1608497	8,1
2005	1798630	9,2
2006	1896515	8,3
2007	1780627	8,6
2008	1283374	13,9
2009	1082587	18,83
2010	961601	20,05

a) Existeix cap tipus de relació lineal entre les variables? És forta aquesta relació? Raoneu les respostes.
b) Calculeu la recta de regressió lineal que relaciona el nombre hipoteques signades en funció de la taxa d'atur.

Exercici 16

La taula següent mostra el nombre d'hores extraordinàries totals en milers (remunerades i no remunerades) realitzades en el conjunt de l'Estat espanyol, així com les taxes d'atur des del primer trimestre del 2008 fins al darrer trimeste de l'any 2010. Les dades han estat extretes de l'INE.

Trimestres	*Nombre total d'hores extres*	*Taxa d'atur*
2010TIV	5574,9	20,33
2010TIII	5058,9	19,79
2010TII	6002,7	20,09
2010TI	6154,1	20,05
2009TIV	6493,2	18,83
2009TIII	6069	17,93
2009TII	7042	17,92
2008TIV	8398,4	13,91
2008TIII	8813,2	11,33
2008TII	9794,4	10,44
2008TI	10058,1	9,63

a) Trobeu, si és pertinent, la recta de regressió que explica el nombre d'hores extres en funció de la taxa d'atur.

b) En el primer trimestre del 2009 la taxa d'atur era del 17,36%. Feu una estimació del nombre d'hores extres en aquest trimestre, així com una mesura de la seua fiabilitat.

Exercici 17

El nombre total d'expedients de regulació del treball al llarg dels anys 2001-2010 a Catalunya i la Comunitat Valenciana, extrets del Ministeri de Treball i Immigració, són els que es mostren en la taula, a excepció de les dades de l'any 2005 que s'han omés.

ANY	*Catalunya*	*Comunitat Valenciana*
2001	661	465
2002	724	494
2003	608	594
2004	565	619
2006	455	413
2007	470	487
2008	874	1286
2009	3964	3490
2010	3318	2810

Se sap que el nombre d'expedients l'any 2005 a Catalunya va ser de 512. Feu una estimació, si convé, del nombre d'expedients en la Comunitat Valenciana, així com una mesura de l'ajust.

Exercici 18

En un museu es vol estudiar la repercussió que tenen les queixes realitzades pels visitants i els ingressos. Per a realitzar-ho, s'observaren les dues variables al llarg de les últimes deu setmanes. Les visites estan expressades en desenes d'assistents.

Queixes	18	26	30	33	38	39	42	44	46	49
Visites	107	105,5	105	104,4	104,3	104	103,7	103,4	103,1	103

Si l'entrada al museu té un cost de 3,6 euros, estimeu els ingressos del museu si en una setmana s'hagueren produït 43 queixes.

Exercici 19

La taula següent mostra el nombre de persones ocupades distribuïdes atenent el sou net de l'activitat principal que desenvolupen (en centenes d'euro) i el nivell d'estudis que tenien l'any 2010, segons dades recollides del Ministeri de Treball i Immigració. Cal dir però, que la variable nivell d'estudis ha estat convertida a numèrica discreta per ser graduable. Així l'equivalència és*: 1 = menys que primaris; 2 = primaris; 3 = secundaris; 4 = batxillerat; 5 = formació professional i 6 = universitaris.* Aquesta conversió s'ha fet a efectes didàctics.

Nivell d'estudis \ SOU	[0, 6)	[6,10)	[10,12)	[12,16)	[16,21)	[21,30)	[30, 40)
1	8,75	21,67	*15,00*	7,93	5,65	*0,74*	1,56
2	293,18	790,92	*601,61*	472,64	92,82	*56,30*	3,77
3	538,08	1551,52	*1226,20*	1098,34	340,72	*74,78*	13,56
4	323,39	670,29	*607,31*	709,62	313,35	*142,00*	53,20
5	303,28	801,87	843,80	982,42	444,90	183,90	50,73
6	164,47	439,26	619,51	1155,75	1230,07	919,40	282,14

a) Estan relacionats linealment el sou i el nivell d'estudis?
b) Calculeu una estimació del sou que cobraria una persona ocupada que tinguera un nivell d'estudis equivalent a 4,5, així com la seua fiabilitat.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar, en cas de ser necessari, a l'hora de realitzar els exercicis i problemes. És convenient no fer-ne un abús excessiu, és a dir, abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10-15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2, i en darrer terme la solució.

Ajudes Tipus 1

Exercici 1

Consulteu la introducció teòrica.

Exercici 2

Consulteu la introducció teòrica. En l'apartat *e*) observeu que demana el percentatge de les persones enquestades que tenen més de 51 anys o uns ingressos de 4000 euros o més. Cal comptar doncs, el nombre de persones que compleixen una condició o l'altra.

Exercici 3

Consulteu la introducció teòrica.

Exercici 4

Consulteu la introducció teòrica.

En l'apartat *c*) heu de tenir present que la variable que condiciona, en aquest cas l'edat, inclou més d'un interval. Llavors, cal agrupar les freqüències conjuntes adequadament.

Exercici 5

Consulteu la introducció teòrica.

En l'apartat *c*) heu de calcular l'estadístic que permet contrastar si dues variables estadístiques estan relacionades linealment.

Exercici 6

Per a fer el gràfic heu de seguir les indicacions de l'exercici 5. Per a afirmar que el gràfic és coherent amb el resultat de la covariància heu de fixar-vos en el signe de l'estadístic.

Per a respondre a l'apartat *c*) cal que apliqueu una propietat.

Exercici 7

Consulteu la introducció teòrica.

Exercici 8

El coeficient de correlació lineal informa de les mateixes coses que la covariància. Podeu consultar la introducció teòrica per a saber-ne les propietats.

Exercici 9

S'ha de calcular un estadístic que mesure el grau de relació lineal entre dues variables.

Exercici 10

El director podria haver eliminat una de les dues proves sempre que les dues discriminen a les mateixes persones.

Exercici 11

L'apartat *a*) és directe i per a la resta d'apartats fa falta construir les funcions lineals que milloren l'ajust. Per a fer les prediccions cal substituir els valors de les variables explicatives en les fórmules.

Exercici 12

Cal fer el mateix que a l'exercici 11, però en aquest cas heu de calcular els estadístics a partir de les dades. Únicament caldrà determinar la recta de regressió que tinga un coeficient de correlació superior a 0,8.

Exercici 13

Vegeu l'ajuda de l'exercici 12.

Exercici 14

Per a contestar a la pregunta *a*) heu de calcular l'estadístic que informa sobre el grau de relació lineal entre dues variables.

Exercici 15

Per a contestar a la pregunta *a*) heu de calcular l'estadístic que informa sobre el grau de relació lineal entre dues variables.

Exercici 16

Únicament quan siga pertinent convindrà calcular la recta i fer-ne la predicció.

Exercici 17

Únicament quan siga pertinent convindrà calcular la recta i fer-ne la predicció.

Exercici 18

Heu de tenir en compte que els ingressos depenen completament del nombre de visites. És a dir, si teniu una estimació de les visites, tindreu una estimació dels ingressos.

Exercici 19

Per a contestar a la pregunta de l'apartat *a*) heu de calcular el coeficient de correlació. Utilitzeu la taula per a fer els càlculs dels estadístics i els productes que es necessiten per a calcular la covariància.

Ajudes Tipus 2

Exercici 1

Per a contestar a la pregunta *b*) cal construir la taula de la distribució marginal, i per a la pregunta *c*), la taula de la distribució condicionada:

Cursos de formació	n_i	$x_i \cdot n_i$
4	1	4
5	3	15
6	2	12
7	7	49
8	6	48
9	5	45
10	1	10
	25	183

X/(Y = 7)	n_i	$x_i \cdot n_i$
6	2	12
7	4	28
8	2	16
	8	**56**

Per a fer els recomptes pots ajudar-te de la taula de doble entrada:

Y = cursos d'idiomes

X = cursos de formació

X \ Y	3	5	6	7	8	10	ni.
4	1						1
5		2	1				3
6				2			2
7				4	3		7
8				2	4		6
9					3	**2**	5
10						**1**	1
n.j	1	2	1	8	10	3	25

Exercici 2

En l'apartat *e*) observeu que demana el percentatge de les persones enquestades que tenen més de 51 anys o uns ingressos de 4000 euros o més. Advertiu que aquest nombre demanat és igual a: (enquestats que ingressen més de 4000 euros) + (enquestats que tenen més de 51 anys) – (enquestats que tenen més de 51 anys i ingressen més de 4000 euros).

Exercici 3

Com que és molt similar a l'exercici 1 consulteu-ne les ajudes.

Exercici 4

Molt similar a l'exercici 1. Consulteu les ajudes d'aquest exercici. Calculeu prèviament les marques de les classes de cada variable. A l'apartat *c*) podeu emprar la taula següent:

EDAT	[0, 6)	[6,10)	[10,12)	[12,16)	[16,21)	[21,30)	[30,40)
[30,45)	566,18	1777,07	*1671,91*	2190,02	1248,87	*736,77*	155,06
[45,55)	323,65	797,11	*881,81*	1123,69	724,93	*448,78*	138,99
[55,65)	185,2	430,59	*503,77*	568,69	306,2	*225,13*	123,53
n.j	**1075,03**	**3004,77**	**3057,49**	**3882,4**	**2280**	**1410,68**	**417,58**

Exercici 5

En l'apartat *c*) heu de calcular la covariància.

Exercici 6

Per a respondre a l'apartat *c*) heu d'aplicar la propietat de la covariància, tenint en compte que cal definir dues variables noves a partir de les dues anteriors: X' = despeses en el segon període i Y'= nombre de passatgers en el segon període. Segons l'enunciat X'=1,1 · X i Y'=1,08 · Y.

Exercici 7

Mateixa ajuda que a l'exercici 6. En aquest cas les noves variables són: X' = nota del test 1 després de l'augment i Y' = nota del test 2 després de l'augment. Segons l'enunciat X' = 1,05 · X i Y' = 1,058 · Y.

Exercici 8

Es remet a l'ajuda de tipus 1 per ser prou aclaridora.

Exercici 9

Cal determinar el coeficient de correlació lineal.

Exercici 10

El director podria haver eliminat una de les dues proves sempre que existisca una relació quasi funcional entre les dues variables.

Exercici 11

Les funcions lineals que s'han de calcular són les rectes de regressió. Per a saber la fiabilitat de les prediccions cerqueu el coeficient de determinació.

Exercici 12

Es remet a l'ajuda de tipus 1 per ser prou aclaridora.

Exercici 13

Vegeu l'ajuda de l'exercici 12

Exercici 14

Per a contestar a la pregunta de l'apartat *a*) heu de calcular el coeficient de correlació.

Exercici 15

Per a contestar a la pregunta de l'apartat *a*) heu de calcular el coeficient de correlació.

Exercici 16

Únicament si el coeficient de correlació és pròxim a 1 o a –1 és pertinent calcular la recta i fer-ne la predicció.

Exercici 17

Únicament si el coeficient de correlació és pròxim a 1 o a –1 és pertinent calcular la recta i fer-ne la predicció.

Exercici 18

Si existeix cap tipus de relació lineal entre el nombre de queixes i el nombre de visites, aleshores podrem trobar la recta de regressió entre el nombre de visites i el de queixes i, amb posterioritat, es podran estimar els ingressos.

Exercici 19

Es remet a l'ajuda de tipus 1 per ser prou aclaridora.

Solucions completes

Exercici 1

Una empresa ha entrevistat vint-i-cinc dels seus treballadors amb tasques administratives per a conèixer el grau d'implicació en la seua formació professional. A cadascun li preguntà el nombre de cursos de formació de més de 30 hores i el nombre de cursos de perfeccionament d'idiomes que havia realitzat en els darrers tres anys. Els resultats són els que es mostren en la taula següent:

Formació ***EST***	8	9	4	5	6	7	7	9	10	7	5	6	7
Idiomes ***ECO***	8	8	3	5	7	7	8	10	10	7	6	7	8

Formació ***EST***	8	5	8	9	8	8	7	7	9	9	8	7
Idiomes ***ECO***	7	5	8	8	7	8	7	7	8	10	8	8

Solució

a) Construïu la taula de freqüències conjunta.

Per a construir la taula de doble entrada cal, en primer lloc, ordenar les dades de les dues variables de menor a major i construir una graella, de manera que en la primera fila se situen les diferents categories o valors que pren una de les variables, i en la primera columna els valor o les categories relatives a la segon. Així, el nombre que cal associar a cada cel·la interior de la taula de doble entrada és la freqüència absoluta conjunta de la dada bivariant formada pels valors corresponents ubicats a la primera fila i a la primera columna.

En el nostre cas podem representar en les files la variable cursos de formació i en les columnes la variable cursos d'idiomes. Així doncs:

M. J. Beltrán / P. J. Verdoy / M. J. Peris - ISBN: 978-84-695-7233-7

Problemes resolts d'estadística aplicada - UJI - DOI: http://dx.doi.org/10.6035/Sapientia72

X = Cursos de formació \ Y = Cursos d'idiomes	3	5	6	7	8	10	ni.
4	1						1
5		2	1				3
6				2			2
7				4	3		7
8				2	4		6
9					3	2	5
10						1	**1**
n.j	**1**	**2**	**1**	**8**	**10**	**3**	**25**

b) Calculeu el nombre mitjà de cursos formació i el nombre mitjà de cursos d'idiomes que han realitzat els treballadors de l'empresa.

Per a contestar aquestes preguntes cal estudiar les dues distribucions marginals; X = cursos de formació i Y = cursos d'idiomes. És a dir:

Cursos de formació	n_i	$x_i \cdot n_i$
4	1	4
5	3	15
6	2	12
7	7	49
8	6	48
9	5	45
10	1	10
	25	183

Cursos d'idiomes	n_i	$y_i \cdot n_i$
3	1	3
5	2	10
6	1	6
7	8	56
8	10	80
10	3	30
	25	185

$$\bar{x} = \frac{183}{25} = 7{,}32 \text{ cursos} \qquad \bar{y} = \frac{185}{25} = 7{,}4 \text{ cursos}$$

c) Calculeu el nombre mitjà de cursos de formació que han fet aquells treballadors que n'han fet set de perfeccionament dels idiomes.

En aquest apartat cal, en primer lloc, construir la variable cursos de formació condicionada a què el valor de la variable cursos d'idiomes és 7. És a dir, cal construir la variable X/(Y = 7). Extraient de la taula aquest distribució marginal:

X/(Y = 7)	n_i	$x_i \cdot n_i$
6	2	12
7	4	28
8	2	16
	8	**56**

En conseqüència:

$$\overline{X/(Y=7)} = \frac{56}{8} = 7 \text{ cursos}$$

d) Quina proporció de treballadors ha realitzat més de cinc cursos en ambdues categories? Quina proporció de treballadors ha fet més de cinc cursos de formació? I més de cinc cursos d'idiomes?

Y = Cursos d'idiomes

X = Cursos de formació	3	5	6	7	8	10	ni.
4	1						1
5		2	1				3
6				2			2
7				4	3		7
8				2	4		6
9					3	2	5
10						1	1
n.j	1	2	1	8	10	3	25

Per a contestar a la primera pregunta, cal que comptem el nombre de dades que compleixen les dues condicions al mateix temps, és a dir, les dades que apareixen amb la cel·la de color vermell:

Així doncs, com que n'hi ha 17, el percentatge demanat és $\frac{17}{25} = \frac{0,68}{1} = \frac{68}{100}$, un 68%.

De la mateixa manera, es compta el nombre de persones que han fet més de 5 cursos de formació, que en són 17, la qual cosa representa un 68%. També podem calcular fàcilment el nombre de treballadors que han fet més de 5 cursos d'idiomes, que en són 18; els quals representen un 72%.

e) Quina proporció de treballadors ha realitzat més de set cursos de formació i més de 8, d'idiomes?

En aquest apartat cal comptar el nombre de treballadors que compleixen totes dues condicions. En aquest cas en són 3 (les cel·les que apareixen en groc a l'apartat *d*)). Per tant, el percentatge és de 3 de 25, un 12%.

f) Quin percentatge dels treballadors que han fet cinc cursos o més de formació, han fet set o més cursos d'idiomes?

En aquest apartat cal observar que canvia la quantitat total sobre la qual hem de fer el percentatge. És a dir, no se'ns demana el percentatge sobre les 25 dades, sinó sobre les que han fet cinc cursos o més de formació, que en són 24. D'aquestes, n'hi ha 21 que han fet 7 cursos o més d'idiomes. En conseqüència, n'hi ha un 87,5%.

Exercici 2

Una empresa vol obrir un punt de venda en un barri d'una gran ciutat de la Comunitat Valenciana. Com que el segment de població a qui va adreçat el producte és per a persones d'edats compreses entre 45 i 55 anys, ha decidit enquestar a una mostra de 50 veïns d'edat compresa en aquesta franja del barri. La taula següent exposa dues de les preguntes que apareixien a l'enquesta: edat i ingressos mensuals en milers d'euros.

Edat	50	51	53	50	51	48	50	49	52	52	49	50	52	51	52	49	50
Ingressos mensuals	3.2	4.1	4.5	3	3.6	2.9	3.8	3.8	3.6	3.9	3	3.8	4.1	3.5	4.0	3.1	3.1

Edat	51	50	51	52	53	52	52	51	50	51	54	50	51	51	51	52	51
Ingressos mensuals	4.3	3.3	3.9	3.7	4.1	4.2	3.5	3.8	3.6	3.4	4.6	3.5	3.6	3.1	4	3.8	4.2

Edat	52	51	50	51	49	51	48	50	52	53	52	50	52	51	51	51
Ingressos mensuals	4	4.4	3.9	3.7	3.4	3.3	2.7	3.4	3.6	4.4	4.3	3.3	4.2	4.2	3.3	3.7

Solució

a) Construïu la taula de doble entrada agrupant els ingressos mensuals en intervals d'amplària 0,5 i de manera que l'extrem petit de la primera classe siga 2,5.

En primer lloc, cal construir els intervals de la variable que ha d'estar agrupada; en aquest cas els ingressos mensuals. Així doncs, seguint les indicacions que dóna l'enunciat els intervals són [2,5 , 3); [2,5 , 3); [3 , 3,5); [3,5 , 4,); [4, 4,5); [4,5 , 5).

Seguint ara les indicacions de l'exercici anterior, apartat *a*), podem construir la taula de doble entrada amb les variables X = ingressos mensuals i Y = edat.

	Y = Edat							
X = Ingressos mensuals	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**ni.**
[2,5 , 3)	2							**2**
[3, 3,5)		3	6	4				**13**
[3,5 , 4)		1	5	7	6			**19**
[4 , 4,5)				6	6	2		**14**
[4,5 , 5)						1	1	**2**
n.j	**2**	**4**	**11**	**17**	**12**	**3**	**1**	**50**

b) Quins són els ingressos mitjans dels enquestats que tenen 51 anys? Quin percentatge té uns ingressos inferiors a 4000 euros?

Cal construir la variable ingressos condicionada a què l'edat siga de 51 anys. És a dir X/Y = 51. A més, com que la variable X està agrupada en intervals és necessari calcular les marques de classe per a obtenir la mitjana demanada. Així, la taula de freqüències d'aquesta variable és:

X/Y = 51	c_i	n_i	$c_i \cdot n_i$
[3, 3,5)	3,25	4	13
[3,5 , 4)	3,75	7	26,25
[4 , 4,5)	4,25	6	25,5
n.j		17	64,75

En conseqüència, la mitjana serà:

$$\overline{X/(Y=51)} = \frac{64,75}{17} = 3,809 \text{ milers €}$$

D'altra banda, hi ha 11 enquestats amb un sou inferior a 4000 euros, la qual cosa representa un 64,7%.

c) Quina és la mitjana d'edat dels enquestats que tenen uns ingressos entre 3500 i 4000 euros? Quin percentatge d'aquests té 50 o 51 anys?

Ara cal construir la variable edat condicionada a uns ingressos d'entre 3500 i 4000 euros. És a dir, Y/(X = [3,5 , 4)). La taula de freqüències de la variable és:

Y/X=[3,5 , 4)	n_i	$x_i \cdot n_i$
49	1	49
50	5	250
51	7	357
52	6	312
n.j	19	968

En conseqüència, la mitjana serà:

$$\overline{Y/(X=[3,5\,,4)} = \frac{968}{19} = 50,95 \text{ anys}$$

D'altra banda, n'hi ha 12 que tenen 50 o 51 anys, la qual cosa representa un 63,16%.

d) Quin percentatge del enquestats ingressen mensualment 4000 o més euros i tenen més de 50 anys?

Cal fer el recompte de les persones enquestades que compleixen les dues condicions que cita l'enunciat, les quals es tradueixen en què han d'ingressar més de 4000 euros (X>4) i tenir més de 50 anys (Y>50).

Revisant la taula s'observa que n'hi ha 16, els quals representen el 32% dels 50 enquestats.

e) Quin percentatge de les persones enquestades tenen més de 51 anys o uns ingressos de 4000 euros o més?

La pregunta és diferent a l'anterior, ja que ens demana quin percentatge compleix una condició o l'altra. És a dir, cal comptar els enquestats que ingressen més de 4000 euros (cel·les amb major grandària) o tenen més de 51 anys (cel·les vermelles). Cal notar que no podem sumar el nombre de persones que compleixen una condició més el nombre de persones que compleixen l'altra, ja que d'aquesta manera estaríem comptant les persones que compleixen ambdues condicions dues vegades.

Y = Edat

X = Ingressos mensuals	48	49	50	51	52	53	54	ni.
[2,5 , 3)	2							2
[3, 3,5)		3	6	4				13
[3,5 , 4)		1	5	7	6			19
[4 , 4,5)				**6**	**6**	**2**		14
[4,5 , 5)						**1**	**1**	2
n.j	2	4	11	17	12	3	1	50

Per tant, el nombre d'enquestats demanats són: (enquestats que ingressen més de 4000 euros) + (enquestats que tenen més de 51 anys) – (enquestats que tenen més de 51 anys i ingressen més de 4000 euros) = 16 + 16 – 10 = 22. Per tant, el percentatge que representen respecte a 50 és del 44%.

Exercici 3

El departament de recursos humans d'una empresa ha decidit realitzar dos tests per tal de seleccionar les persones que hauran de fer-se càrrec d'un projecte d'innovació. Les notes obtingudes pels aspirants es mostren en la taula següent:

TEST 1	7	6	5	4	5	8	7	8	9	6	5	8	6	8	7	8	7	6	6	9
TEST 2	8	7	6	6	7	10	9	9	10	8	6	10	8	9	8	8	7	8	6	8

Solució

a) Construir la taula de doble entrada.

De la mateixa manera que fèiem en l'apartat *a*) del primer exercici, construïm la taula de doble entrada.

Y = Test 1

X = Test 2	**4**	**5**	**6**	**7**	**8**	**9**	**ni.**
6	1	2	1				**4**
7		1	1	1			**3**
8			3	2	1	1	**7**
9				1	2	0	**3**
10					2	1	**3**
n.j	**1**	**3**	**5**	**4**	**5**	**2**	**20**

b) Calculeu la nota mitjana en el test 2 dels aspirants que han obtingut un 6 en el test 1.

Se'ns demana que calculem la mitjana aritmètica de la variable test 2 condicionada a què la variable test 1 és 6. És a dir, X/Y = 6. Calculem la taula de freqüències corresponent.

Y/X = 6	**ni**	**xi · ni**
6	1	6
7	1	7
8	3	24
n.j	5	37

En conseqüència, la mitjana serà:

$$\overline{X/(Y=6)} = \frac{37}{5} = 7{,}4 \text{ punts}$$

c) Percentatge d'aspirants que obtenen un nota inferior a 8 en el test 2 d'entre aquells que obtenen una nota superior a 6 en el test 1.

De la mateixa manera que a l'apartat *f*) de l'exercici 1, en aquest apartat no se'ns demana el percentatge sobre els 20 aspirants, sinó sobre els que han tret més de 6 en el test 1, que en són 11. D'aquestes n'hi ha 1 que ha tret menys de 8. En conseqüència, n'hi ha un 9,09%.

Exercici 4

La taula següent mostra el nombre de persones ocupades distribuïdes atenent el sou net de l'activitat principal que desenvolupen (en centenes d'euro) i l'edat l'any 2010, segons dades recollides del Ministeri de Treball i Immigració.

	SOU						
EDAT	[0, 6)	[6,10)	***[10,12)***	***[12,16)***	***[16,21)***	***[21,30)***	***[30, 40)***[6]
[16,25)	289,79	490,44	*249,08*	126,47	38,03	*1,70*	0
[25,30)	232,55	673,68	*571,85*	430,16	192,86	*20,80*	11,01
[30,45)	566,18	1777,07	*1671,91*	2190,02	1248,87	*736,77*	155,06
[45,55)	323,65	797,11	*881,81*	1123,69	724,93	*448,78*	138,99
[55,65[7])	185,20	430,59	*503,77*	568,69	306,20	*225,13*	123,53

Solució

a) Construïu les taules de freqüència de les distribucions de les variables marginals i calculeu la mitjana aritmètica de cadascuna.

De la mateixa manera que a l'apartat *a*) de l'exercici 1, es construeixen les taules de freqüència de les variables marginals X = edat de les persones ocupades i Y = sou de les persones ocupades.

Edat	c_i	n_i	$c_i \cdot n_i$
[16,25)	20,5	1195,51	24507,955
[25,30)	27,5	2132,91	58655,025
[30,45)	37,5	8345,88	312970,5
[45,55)	50	4438,96	221948
[55,65)	60	2343,11	140586,6
n.j		18456,37	758668,08

Sou	c_i	n_i	$c_i \cdot n_i$
[0, 6)	3	1597,37	4792,11
[6,10)	8	4168,89	33351,12
[10,12)	11	3878,42	42662,62
[12,16)	14	4439,03	62146,42
[16,21)	18,5	2510,89	46451,47
[21,30)	25,5	1433,18	36546,09
[30,40)	35	428,59	15000,65
		18456,37	240950,5

6. En la taula original el darrer interval és 3000 euros o més. S'ha tancat l'interval per a fer l'exercici.
7. En la taula original el darrer interval és 55 anys o més. S'ha tancat l'interval per a fer l'exercici.

I calculem també les mitjanes:

$$\bar{x} = \frac{758668,08}{18456,37} = 41,12 \text{ anys} \qquad \bar{y} = \frac{240950,5}{18456,37} = 13,06 \text{ milers €}$$

Cal notar que, encara que l'exercici no ho demana explícitament, és convenient calcular la desviació típica, per a conèixer una mesura de dispersió de les dades.

b) Construïu la taula de freqüència de l'edat d'aquelles persones ocupades que tenen un sou de 1200 a 1600 euros. Calculeu també l'edat mitjana de les persones que cobren entre 1200 i 1600 euros.

Construïm la taula de freqüències de la variable edat condicionada a què la variable sou estiga compresa entre 1200 i 1600 euros. És a dir, X/(Y=[1200,1600)). Així doncs:

X/(Y = [1200,1600))	c_i	n_i	$c_i \cdot n_i$
[16,25)	20,5	126,47	2592,635
[25,30)	27,5	430,16	11829,4
[30,45)	37,5	2190,02	82125,75
[45,55)	50	1123,69	56184,5
[55,65)	60	568,69	34121,4
n.j		4439,03	186853,69

En conseqüència, la mitjana serà:

$$\overline{X/(Y=[12,16)} = \frac{186853,69}{4439,03} = 42,1 \text{ anys}$$

c) Construïu la taula de freqüència del sou d'aquelles persones ocupades que tenen 30 anys o més. Quin sou mitjà cobren?

Cal notar que la variable que condiciona, en aquest cas l'edat, inclou més d'un interval. Llavors cal agrupar les freqüències conjuntes adequadament:

EDAT	***[0, 6)***	***[6,10)***	***[10,12)***	***[12,16)***	***[16,21)***	***[21,30)***	***[30,40)***
[30,45)	566,18	1777,07	*1671,91*	2190,02	1248,87	*736,77*	155,06
[45,55)	323,65	797,11	*881,81*	1123,69	724,93	*448,78*	138,99
[55,65)	185,2	430,59	*503,77*	568,69	306,2	*225,13*	123,53
n.j	**1075,03**	**3004,77**	**3057,49**	**3882,4**	**2280**	**1410,68**	**417,58**

en conseqüència, la variable sou condicionada a què l'edat siga de 30 anys o més és:

Y/(X≥30)	c_i	n_i	$c_i \cdot n_i$
[0, 6)	3	1075,03	3225,09
[6,10)	8	3004,77	24038,16
[10,12)	11	3057,49	33632,39
[12,16)	14	3882,4	54353,6
[16,21)	18,5	2280	42180
[21,30)	25,5	1410,68	35972,34
[30,40)	35	417,58	14615,3
		15127,95	208016,88

I per tant la mitjana aritmètica de la variable demanada és:

$$\overline{Y/(X>30)} = \frac{208016,88}{15127,95} = 13,751 \text{ milers €}$$

d) Quin percentatge de persones ocupades tenen 45 anys o més i cobren 1600 euros o més?

Cal cercar el nombre d'ocupats que compleix les dues variables. Així n'hi ha –observant en la taula i sumant els nombres adequats– 1967,56 milers, la qual cosa representa un 10,66%.

e) Quin percentatge de persones ocupades tenen 45 anys o més o cobren 1600 euros o més?

De la mateixa manera que a l'apartat *f*) de l'exercici 2, cal comptar els ocupats que tenen 45 anys o més (cel·les amb major grandària) o cobren 1600 euros o més (cel·les vermelles).

SOU

EDAT	[0, 6)	[6,10)	[10,12)	[12,16)	[16,21)	[21,30)	*[30,40)*	**ni.**
[16,25)	289,79	490,44	*249,08*	126,47	**38,03**	***1,7***	**0**	**1195,51**
[25,30)	232,55	673,68	*571,85*	430,16	**192,86**	***20,8***	**11,01**	**2132,91**
[30,45)	566,18	1777,07	*1671,91*	2190,02	**1248,9**	***736,77***	**155,06**	**8345,88**
[45,55)	323,65	797,11	*881,81*	1123,69	**724,93**	***448,78***	**138,99**	**4438,96**
[55,65)	185,2	430,59	*503,77*	568,69	**306,2**	***225,13***	**123,53**	**2343,11**
n.j	**1597,37**	**4168,89**	**3878,42**	**4439,03**	**2510,89**	**1433,18**	**428,59**	**18456,37**

De la mateixa manera que a l'exercici 2, el nombre d'ocupats que complixen les dues condicions són:

(45 anys o més) + (1600 euros o més) – (45 anys o més i 1600 euros o més)

6782,07 + 4372,66 – 1967,56 = 9187,17

Per tant, el percentatge demanat és del 49,788%.

f) Quin percentatge d'ocupats tenen menys de 30 anys dels que cobren 1200 euros o més?

En primer lloc, cal saber quants ocupats cobren 1200 euros o més. Sumant els valors de la taula s'obtenen 8.811,69 milers de persones (requadre roig). D'aquest n'hi ha 821,03 (ombrejat groc) que tenen menys de 30 anys. Per tant, el percentatge demanat és 9,32%.

	SOU						
EDAT	***[0, 6)***	***[6,10)***	***[10,12)***	***[12,16)***	***[16,21)***	***[21,30)***	***[30,40)***
[16,25)	*289,79*	*490,44*	*249,08*	***126,47***	***38,03***	***1,7***	***0***
[25,30)	*232,55*	*673,68*	*571,85*	***430,16***	***192,86***	***20,8***	***11,01***
[30,45)	*566,18*	*1777,07*	*1671,91*	*2190*	*1248,87*	*736,77*	*155,06*
[45,55)	*323,65*	*797,11*	*881,81*	*1123,7*	*724,93*	*448,78*	*138,99*
[55,65)	*185,2*	*430,59*	*503,77*	*568,69*	*306,2*	*225,13*	*123,53*

Exercici 5

La taula següent mostra la població en edat de treballar analfabeta en la Comunitat Valenciana, Madrid, Andalusia i el País Basc al llarg dels anys 2000-2010 en milers de persones.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Andalusia	332,28	280,99	342,14	307,41	294,91	247,90	262,25	278,03	294,34	290,37	279,91
Madrid	73,15	58,88	74,06	83,33	71,83	47,35	43,35	40,34	46,68	56,98	72,21
País Basc	17,69	13,77	14,67	12,56	13,27	9,37	11,34	12,50	13,66	14,23	11,10
C. Valenciana	128,01	96,83	110,81	117,30	114,73	69,46	79,91	79,01	92,45	99,06	85,76

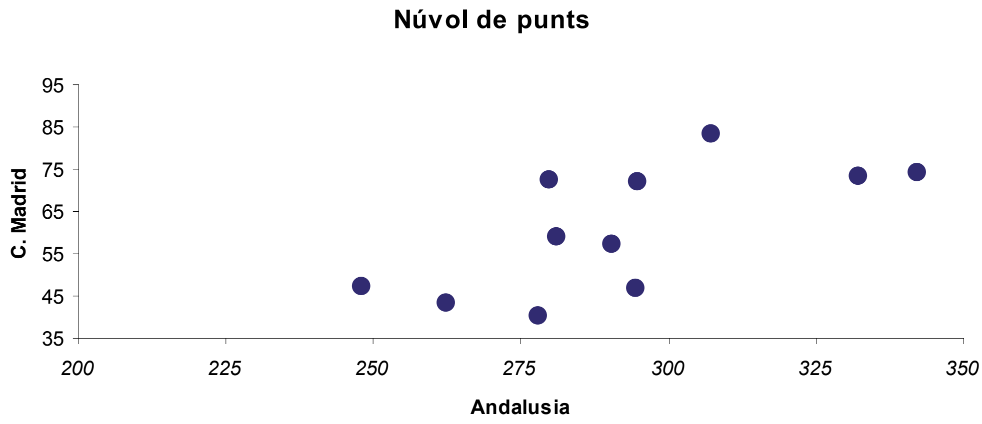
Font: INE

Solució

a) Representeu el núvol de punts entre les variables: població en edat de treballar a Andalusia i població en edat de treballar a la Comunitat Valenciana. Què observeu pel que fa a la existència o no de la relació lineal entre les dues variables?

Un núvol de punts és un gràfic de dues dimensions en el qual es representen els valors de les dues variables. Cada punt del núvol té coordenada x (abscissa o horitzontal) el valor d'una de les variables, i coordenada y (vertical o ordenada) el valor que li correspon de l'altra variable. La forma d'aquest gràfic és el primer pas per a saber si dues variables estan correlacionades.

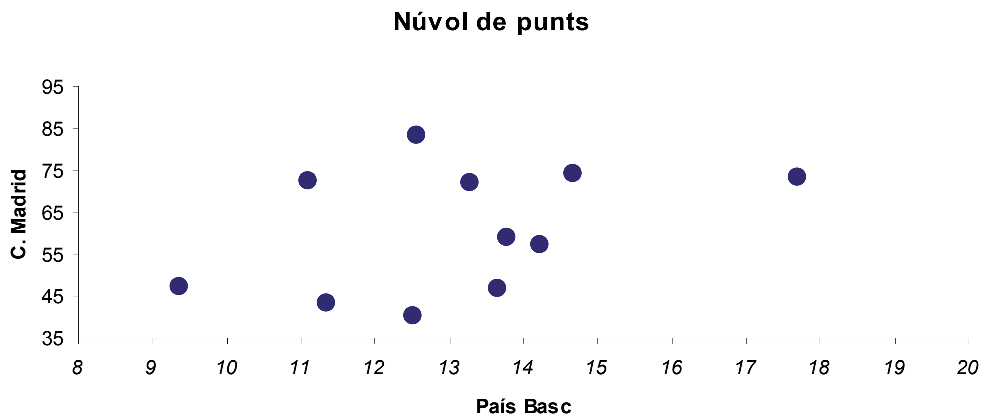
En el nostre cas podem representar en l'eix de les abscisses la població activa i analfabeta a Andalusia, i en l'eix d'ordenades les persones analfabetes i actives de Madrid. Així quda el gràfic:



Pel que s'observa visualment, sembla haver-hi una relació lineal positiva entre les dues variables.

b) Representeu el núvol de punts entre les variables: població en edat de treballar en Madrid i població en edat de treballar al País Basc. Què observeu pel que fa a l'existència o no de la relació lineal entre les dues variables?

De la mateixa manera que a l'apartat anterior, el gràfic queda així:



S'hi observa que també existeix una relació lineal positiva, encara que en aquest cas no sembla tan clar perquè el núvol de punts és més «ample». És a dir, sembla que les dades no seguisquen un línia recta creixent amb tanta claredat com a l'apartat anterior.

c) Calculeu l'estadístic adequat per a confirmar les suposicions que has fet en els dos apartats anteriors.

En els dos apartats anteriors hem observat a partir del gràfic que existeix una relació lineal positiva entre les dues variables. L'estadístic que permet contrastar aquesta hipòtesi és la covariància. Si aquest estadístic té signe positiu llavors existeix relació lineal entre les dues variables i aquesta és positiva. Si per contra té signe negatiu, llavors també existeix relació lineal però en aquest cas és negativa. Si la covariància és zero, aleshores les dues variables no tenen relació lineal.

L'expressió de la covariància és : $S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{(x_i - \overline{X})(y_j - \overline{Y})\cdot n_{ij}}{n}$. No obstant, per a realitzar els problemes emprarem l'expressió equivalent:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y}$$

Així doncs, cal avaluar les covariàncies de tots dos parells de variables.

Andalusia i Madrid

Podem considerar la variable X = actius i analfabets a Andalusia i Y= actius i analfabets a Madrid.

Com s'observa en la fórmula, en primer lloc s'ha de calcular per a cada variable les seues mitjanes aritmètiques. Fent aquests càlculs de la mateixa manera que en la unitat 1 s'obtenen els valors: $\overline{X} = 291{,}866$ i $\overline{Y} = 60{,}742$.

En segon lloc hem de calcular els sumatoris $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Com s'observa, el que cal és multiplicar cada valor de la variable X pel seu corresponent de la variable Y i per la seua freqüència conjunta. Després, cal sumar tots aquests productes i dividir-los entre el nombre total de dades.

Com que en el nostre cas la freqüència conjunta de cada dada bivariant és 1, tan sols cal fer els productes de cada valor d'una variable pel seu corresponent i després fer la suma. Així:

X	332,28	280,99	342,14	307,41	294,91	247,9	262,25	278,03	294,34	290,37	279,91
Y	73,15	58,88	74,06	83,33	71,83	47,35	43,35	40,34	46,68	56,98	72,21
$x_i \cdot y_j$	**24306,28**	**16544,69**	**25338,89**	**25616,48**	**21183,39**	**11738,07**	**11368,54**	**11215,73**	**13739,79**	**16545,28**	**20212,30**

Així, substituint l'expressió anterior:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}=\frac{24306{,}28+16544{,}69+25338{,}89+25616{,}48+\ldots+16545{,}28+20212{,}30}{11}$$

$$=\frac{197809{,}430}{11}=17982{,}675$$

I per tant:

$$S_{XY}=\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}-\overline{X}\cdot\overline{Y}=17982{,}675-291{,}866\cdot 60{,}742=332{,}954.$$

Com que la covariància ens queda positiva, existeix relació lineal positiva. Es conclou, doncs, que el gràfic i l'estadístic estan en concordança.

País Basc i Madrid

Cal fer exactament el mateix, però ara considerant X= actius i analfabets al País Basc i per Y= actius i analfabets a Madrid. Calcular les mitjanes per a cada variable: $\overline{X}=13{,}105$ i $\overline{Y}=60{,}742$.

Calculem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$ tal com hem fet abans:

X	17,69	13,77	14,67	12,56	13,27	9,37	11,34	12,5	13,66	14,23	11,1
Y	73,15	58,88	74,06	83,33	71,83	47,35	43,35	40,34	46,68	56,98	72,21
$x_i \cdot y_j$	**1294,024**	**810,778**	**1086,460**	**1046,625**	**953,184**	**443,670**	**491,589**	**504,250**	**637,649**	**810,825**	**801,53**

Així, substituint l'expressió anterior:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}=\frac{1294{,}024+810{,}778+1086{,}460+1046{,}625+\ldots+810{,}825+801{,}53}{11}$$

$$=\frac{8880{,}594}{11}=807{,}236$$

I per tant:

$$S_{XY}=\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}-\overline{X}\cdot\overline{Y}=807{,}236-13{,}105\cdot 60{,}742=11{,}212.$$

Com que la covariància ens queda positiva, també existeix relació lineal positiva. Es conclou, doncs, que el gràfic i l'estadístic estan en concordança.

Cal notar però, que amb la covariància encara no podem saber el grau de la relació lineal entre totes dues variables. És precís calcular un altre estadístic per a estudiar-lo: el coeficient de correlació lineal.

Exercici 6

Es van recol·lectar els valors mensuals de les despeses en publicitat d'una companyia ferroviària i el nombre de passatgers al llarg de 15 mesos. Les dades les mostra la taula:

Publicitat (en milers)	10	12	8	17	10	15	10	14	19	10	11	13	16	10	12
Passatgers (en milers)	15	17	13	23	16	21	14	20	24	17	16	18	23	15	16

Solució

a) Calculeu la despesa mitjana i el nombre mitjà de passatgers.

Per a calcular la despesa mitjana i el nombre mitjà de passatgers cal estudiar les distribucions marginals de les dues variables. Si anomenem X = despeses en publicitat i Y = nombre de passatgers llavors, les distribucions marginals són:

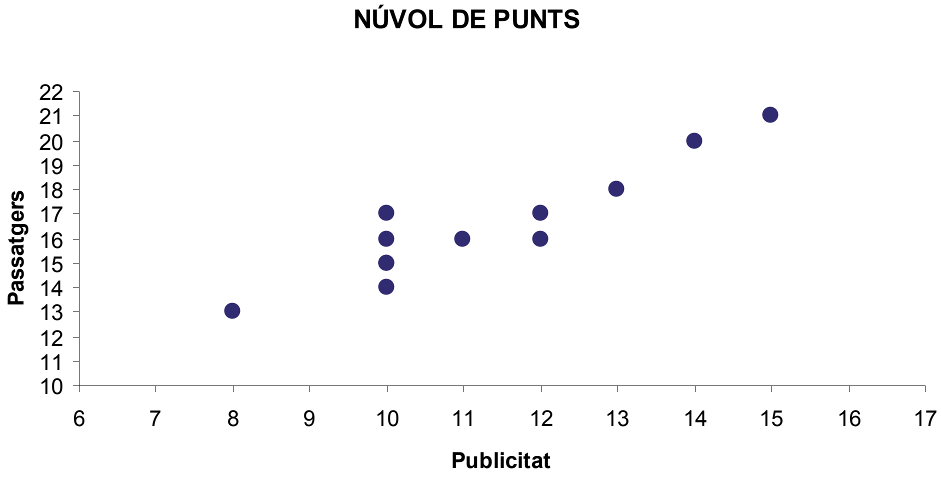
X	ni
8	1
10	5
11	1
12	2
13	1
14	1
15	1
16	1
17	1
19	1
n.j	**15**

Y	**ni**
13	1
14	1
15	2
16	3
17	2
18	1
20	1
21	1
23	2
24	1
n.j	15

De la mateix manera que es fa en la unitat 1, es calculen les mitjanes aritmètiques, que en aquest cas són:

$\overline{X} = 12{,}467$ i $\overline{Y} = 17{,}867$.

b) Feu el núvol de punts i calculeu-ne la covariància. És coherent el valor de l'estadístic amb el núvol de punts?



Per a determinar la covariància cal fer el mateix que a l'exercici 5. Com que les mitjanes ja les hem calculades a l'apartat anterior, ara cal trobar els altres elements components de l'expressió de la covariància. Així:

X	10	12	8	17	10	15	10	14	19	10	11	13	16	10	12
Y	15	17	13	23	16	21	14	20	24	17	16	18	23	15	16
$x_i \cdot y_j$	**150**	204	104	391	160	315	140	280	456	170	176	234	368	150	192

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{150+204+.......+150+192}{15} = 233$$ I per tant, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 233 - 12{,}467 \cdot 17{,}867 = 10{,}252$$

Com es comprova, el núvol de punts concorda amb el valor de la variància.

c) Si per als 15 mesos posteriors es preveu que la inversió en publicitat de cada mes augmente un 10% respecte al mateix mes del període anterior, i també es preveu que aquest fet provocará un augment del 8% en el nombre de passatgers cada mes. Quina serà la covariància en aquest segon període?

Se'ns demana el valor de la covariància de les variables despesa en publicitat i nombre de passatgers en aquest segon període; en el qual les despeses han augmentat un 10% i el nombre de passatgers un 8%. Definim, per tant, aquestes dues noves variables: X' = despeses en el segon període i Y' = nombre de passatgers en el segon període.

Així, segons l'enunciat X'= 1,1 · X i Y' = 1,08 · Y. Per a calcular la covariància entre X' i Y' únicament cal aplicar les propietats.[8]

Aleshores: $S_{X'Y'} = 1{,}1 \cdot 1{,}08\ S_{XY} = 1{,}1 \cdot 1{,}08 \cdot 10{,}252 = 12{,}179$

8. Si tots els valors d'una variable X es multipliquen per una constant *a* i tots els valors de la variable Y per una constant *b*, la covariància queda multiplicada pel producte de les constants. És a dir: = a · b S_{XY}

Exercici 7

Una empresa ha realitzat dos tests psicotècnics als 9 treballadors d'un departament com a part del procés de selecció del nou director del departament. La taula següent mostra els resultats obtinguts pels aspirants:

TEST 1	5	7	6	9	3	1	2	4	6
TEST 2	6	5	8	6	4	2	1	3	7

Solució

a) Calculeu la covariància. Existeix cap tipus de relació lineal entre les dues variables?

De la mateixa manera que en els problemes anteriors, cal trobar cadascun dels components. Anomenem X = notes del test 1 i per Y = notes del test 2 i obtenim les mitjanes de cada variable: $\overline{X} = 4{,}778$ i $\overline{Y} = 4{,}667$.

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho s'han de calcular els productes corresponents i fer-ne la suma.

X	5	7	6	9	3	1	2	4	6
Y	6	5	8	6	4	2	1	3	7
$x_i \cdot y_j$	30	35	48	54	12	2	2	12	42

I per tant:

$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{30+35+48+.....+2+12+42}{9} = 26$ i conseqüentment la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 26 - 4{,}778 \cdot 4{,}667 = 3{,}701$$

Hi ha hagut un error en una pregunta de cada test i el tribunal decideix augmentar un 5% la puntuació de cada participant. Calcula'n novament la covariància.

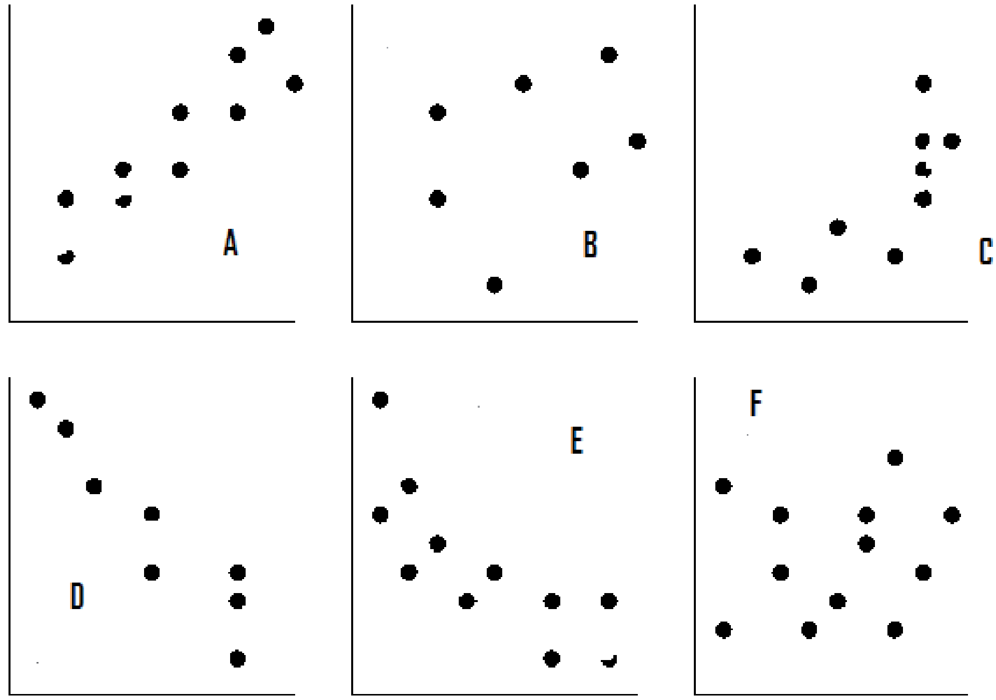
De la mateixa manera que a l'apartat *c*) de l'exercici 6, cal definir unes noves variables i aplicar la mateixa propietat. Definim, per tant, aquestes dues noves variables: X' = nota del test 1 després de l'augment i Y' = nota del test 2 després de l'augment.

Segons l'enunciat X' = 1,05 · X i Y' = 1,058 · Y. Per a calcular la covariància entre X' i Y' únicament cal aplicar-hi les propietats.

Aleshores: $S_{X'Y'} = 1{,}1 \cdot 1{,}08\, S_{XY} = 1{,}1 \cdot 1{,}08 \cdot 3{,}701 = 4{,}397$

Exercici 8

Donats els següents núvols de punts, contesta:



Solució

a) Associeu cada núvol de punts amb el valor del coeficient de correlació que li correspon entre aquests: -0,9; 0,4; 0,95: -0,65; 0,1; 0,6. Raoneu la resposta.

Com ja s'ha comentat en els problemes anteriors d'aquesta unitat, la covariància permet discernir si dues variables X i Y tenen una relació positiva, negativa o zero, però no aporta informació del grau de dependència d'una variable respecte a l'altra. A més a més, la covariància depén de les unitats de mesura emprades per a X i Y –si per exemple X es mesura en m^3 i Y en mm^3, cada desviació de X augmenta S_{XY} 10^9 vegades. Per a fer front a aquestes dues dificultats es definix el concepte de correlació lineal r_{XY} :

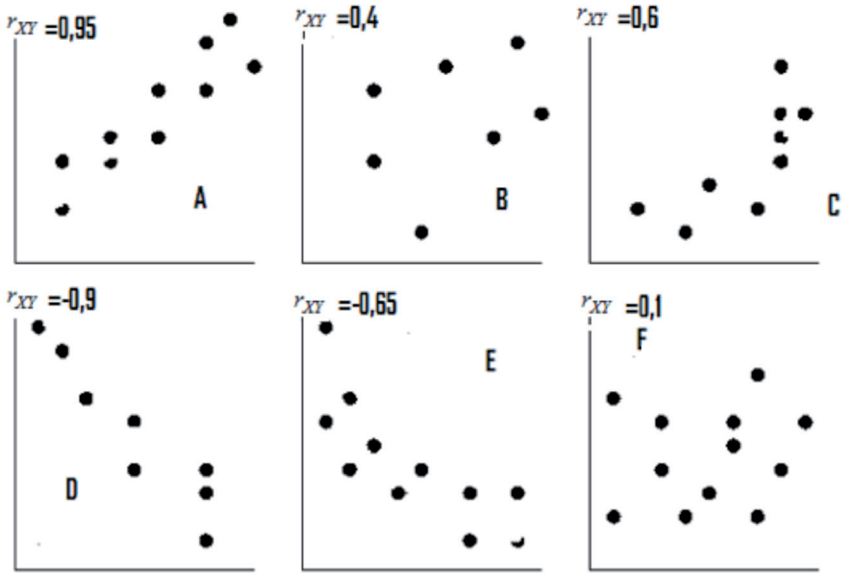
$$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y}$$ sent S_X i S_Y les desviacions típiques de X i Y

És evident, que per definició el coeficient de correlació lineal informa de les mateixes coses que la covariància. A més a més, compleix una propietat molt important, està fitat per 1 i per -1. Així doncs, r_{XY} es caracteriza per:

- Ser adimensional i estar sempre comprés entre –1 i 1.
- Si hi ha relació lineal forta positiva, $r_{XY} > 0$ i està a prop d'1.
- Si hi ha relació lineal negativa forta, $r_{XY} < 0$ i està a prop de –1.
- Si no hi ha relació lineal r_{XY} serà 0.

Així doncs, el coeficient de correlació serà tan pròxim a 1 o –1 com el núvol de punts siga més «estret». Així doncs, a núvols de punts estrets li corresponen valors de r_{XY} r_{XY} pròxims a 1 o –1, i per tant, les dades s'ajustaran bé a una línia recta. Al contrari, valors r_{XY} pròxims al 0, implica que el núvol de punts es més ample i, en conseqüència, les dades no s'ajustaren bé a una línia recta.

En el nostre exercici cal associar cada núvol de punts amb el coeficient de correlació adient. Tenint en compte el que acabem de comentar les associacions seran:



b) Indiqueu per a cada núvol de punts el signe de la covariància i digueu quin n'és el significat.

Tenint present allò que hem dit a l'apartat *a*) el signe del coeficient de correlació és el mateix que el de la covariància. Per tant, els núvols dels gràfics A, B, C i D tindran una covariància positiva (relació lineal positiva) i els núvols E i F tindran una covariància negativa (relació lineal negativa). Cal notar que el gràfic F té una correlació de 0,1. És a dir, malgrat tenir signe positiu la relació lineal de totes dues variables és molt dèbil per ser r_{XY} molt prop de 0.

Exercici 9

La taula següent mostra la despesa total mitjana, la despesa mitjana en aliments i begudes no alcohòliques i la despesa en habitatge, aigua, electricitat, gas i altres combustibles en euros, per nombre de persones que formen la unitat familiar l'any 2009[9], segons dades de l'INE.

Nombre de membres de la llar	*1*	*2*	*3*	*4*	*5*	*6 o més*
Despesa mitjana total	18355,25	27755,08	33414,09	38576,14	40699,09	41562,31
Despesa en habitatge, aigua, electricitat, gas i altres combustibles	7493,88	8990,72	9205,13	9645,19	10114,49	9272,18

Solució

a) Existeix una forta relació lineal entre el nombre de membres que viuen en una llar i la despesa mitjana total? Raoneu la resposta.

Ja hem comentat a l'apartat *a*) de l'exercici 8 que per a conéixer el grau de relació lineal entre dues variables s'ha de calcular el coeficient de correlació. L'expressió és $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y}$. Per tant, hem de trobar la covariància i les desviacions típiques de cada variable.

Anomenem X = nombre de membres del llar i Y = despesa mitjana total, i fem els càlculs necessaris. Les mitjanes i les desviacions típiques les calcularem tal com fèiem a la unitat 1. Així: $\overline{X} = 3{,}667$; $S_X = 1{,}972$ i $\overline{Y} = 33393{,}6$; $S_Y = 8214{,}845$.

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\dfrac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho cal determinar els productes corresponents i fer-ne la suma.

X	*1*	*2*	*3*	*4*	*5*	*7*
Y	18355,25	27755,08	33414,09	38576,14	40699,09	41562,31
$x_i \cdot y_j$	18355,25	55510,16	100242,27	154304,56	203495,45	290936,17

9. Per a fer l'exercici, considera **7** membres l' interval 6 o més membres de la llar.

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{18355,25 + 55510,16 + ... + 203495,45 + 290936,17}{6} = 137140,643$$

i, conseqüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 137140,643 - 3,667 \cdot 33363,93 = 14795,112.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{14795,112}{1,972 \cdot 8214,845} = 0,913$ la qual cosa significa que la relació lineal entre les variables despesa mitjana total i membres que formen una llar és prou forta, qüestió perfectament lògica per altra part.

b) I entre el nombre de membres que viuen en una llar i la despesa en habitatge, aigua, electricitat, gas i altres combustibles? Raona la resposta.

En aquest apartat cal fer exactament el mateix que a l'anterior. Anomenem X = nombre de membres del llar i Y = despesa en habitatge, aigua, electricitat, gas i altres combustibles i fem els càlculs necessaris. $\overline{X} = 3,667$; $S_X = 1,972$ i $\overline{Y} = 9120,265$; $S_Y = 812,016$.

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho, cal determinar els productes corresponents i fer-ne la suma.

X	*1*	*2*	*3*	*4*	*5*	*7*
Y	7493,88	8990,72	9205,13	9645,19	10114,49	9272,18
$x_i \cdot y_j$	7493,88	17981,44	27615,39	38580,76	50572,45	64905,26

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{7493,88 + 17981,44 + ... + 64905,26}{6} = 34524,8633$$

i, conseqüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 34524,863 - 3,667 \cdot 9120,265 = 1080,851.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{1080,851}{1,972 \cdot 812,016} = 0,675$ la qual cosa significa que la relació lineal entre les variables despesa en habitatge, aigua, electricitat, gas i altres combustibles i membres que formen una llar és molt dèbil.

Exercici 10

El director de recursos humans d'una empresa ha realitzat dos tests psicotècnics per a seleccionar les persones que han de treballar en el departament de màrqueting. S'hi han presentat 9 persones i els resultats obtinguts en cadascun dels tests han estat els següents:

TEST 1	175	181	192	211	235	255	275	286	292
TEST 2	169	185	202	219	240	266	295	329	357

Tenint en compte els resultats dels tests, creieu que el director podria haver eliminat un dels dos tests per a decidir els candidats? Raoneu la resposta.

Solució

El director podria haver eliminat una de les dues proves sempre que les dues discriminen a les mateixes persones. És a dir, si existeix una relació quasi funcional entre les dues variables. Així doncs, per a contestar a la pregunta caldrà calcular el coeficient de correlació lineal i, si aquest és pròxim a 1 o –1, llavors la relació lineal entre les dues variables serà forta i, per tant, un test no aportarà informació addicional i es podria eliminar.

Calculem, doncs, r_{XY} sent X = notes test 1 i Y = notes test 2.

Fem ara els càlculs necessaris:

$\overline{X} = 233{,}556$; $S_X = 43{,}169$ i $\overline{Y} = 251{,}333$; $S_Y = 61{,}560$.

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho hem de calcular els productes corresponents i fer-ne la suma.

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{175 \cdot 169 + \ldots\ldots + 292 \cdot 357}{9} = \frac{551746}{9} = 61305{,}111$$

i, consegüentment, la covariància és::

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 61305{,}111 - 233{,}556 \cdot 251{,}333 = 2604{,}781.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{2604{,}781}{43{,}169 \cdot 61{,}560} = 0{,}980$

Com que el coeficient r_{XY} és tan proper a 1, les dues proves permeten escollir les mateixes persones, i per tant una de les dues podria eliminar-se.

Exercici 11

En una mostra de 150 empreses del sector de serveis es recullen dades sobre el nombre de treballadors de l'empresa (X) i la facturació (Y) anual en milions d'euros. Els resultats es mostren resumits en els següents estadístics:

$$\overline{X} = 14\,treballadors;\ \overline{Y} = 100\,milions;\ S_X = 2\,treballadors;\ S_Y = 25\,milions;$$
$$S_{XY} = 45\,treballadors \times milió$$

Solució

a) Calculeu la correlació lineal i interpreta-la.

En aquest cas cal únicament substituir en l'expressió del coeficient de correlació:

$$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{45}{25 \cdot 2} = 0{,}9.$$

Com que està prop d'1, podem dir que la relació lineal és bastant forta.

b) Calcula el model de regressió lineal que millor aproxima la facturació en funció del nombre de treballadors.

La recta de regressió d'una variable Y, anomenada explicada o dependent, respecte a una altra X, anomenada explicativa o independent, és la funció lineal $Y = aX + b$ que millor s'ajusta a les dades emprant el criteri dels mínims quadrats. És a dir, d'una banda cada valor x_i de la distribució de dades té el seu corresponent valor y_i per la distribució de dades. Però a més a més, per a aquest valor x_i també es pot calcular el seu estimat per la recta : $y_i^{'} = ax_i + b$. Doncs bé, el mètode dels mínims quadrats permet obtenir els valors de l'equació de la recta a i b que minimitzen la suma dels quadrats de les distàncies entre y_i i $y_i^{'}$.

Els valors de a i de b que s'obtenen pel mètode dels mínims quadrats depenen òbviament de les dades. Així, els valors són:

$$a = \bar{y} - \frac{s_{xy}}{s_x^2} \cdot \bar{x} \qquad \text{i} \qquad b = \frac{s_{xy}}{s_x^2}$$

I per tant, la recta de regressió de Y sobre X és:

$$Y=\left(\overline{y}-\frac{S_{XY}}{S_X^2}\cdot\overline{x}\right)+\frac{S_{XY}}{S_X^2}\cdot X$$

i recol·locant els termes es pot escriure així:

$$Y-\overline{Y}=\frac{S_{XY}}{S_X^2}\cdot(X-\overline{X}).$$

Si s'haguera pres Y com a variable independent o explicativa, i X com a dependent o explicada, la recta de regressió que *es necessita* seria la que minimitza errors de la X. S'anomena recta de regressió de X sobre Y i es calcula fàcilment permutant els llocs de *x* i *y*, obtenint-se:

$$X-\overline{X}=\frac{S_{XY}}{S_Y^2}\cdot(Y-\overline{Y}).$$

En aquest cas demana la recta que explica la facturació en funció del nombre de treballadors. Per tant, ens demana la recta Y sobre X. Per a calcular-la, tan sols cal fer les substitucions corresponents, ja que l'exercici ens dóna tots els estadístics necessaris.

Substituint:

$y-\overline{Y}=\frac{S_{XY}}{S_X^2}\cdot(x-\overline{X})\rightarrow y-100=\frac{45}{4}\cdot(x-14)\rightarrow$ aïllant la variable *y* mitjançant matemàtiques elementals, obtenim: Y = 11,25 X–57,5.

És a dir:

La facturació = 11,25 · nombre de treballadors – 57,5

a) En funció d'aquest ajust, calculeu de manera aproximada la quantitat que s'espera que facture una empresa amb 15 treballadors. És fiable aquesta predicció? Raoneu la resposta.

Per a fer la predicció únicament cal substituir X per 15, ja que d'aquesta manera obtindrem l'estimació de la facturació per a una empresa que tinguera 15 treballadors. Així doncs:

La facturació = 11,25 · 15 – 57,5 = 111,25 milions.

Per a calcular la fiablilitat cal emprar el *coeficient de determinació lineal (R^2)* el qual es pot definir com el percentatge de variància de Y que es pot explicar per X i, se'l sol anomenar qualitat o bondat de l'ajustament perquè valora la proximitat del núvol de punts a la recta de regressió (o dit d'una altra manera, com està d'ajustat el núvol de punts a la recta de regressió). En les regressions lineals, aquest coeficient té una expressió extremadament simple ja que coincideix amb el quadrat del coeficient de correlació lineal: $r_{XY}{}^2 = R^2$.

Així doncs, en el nostre cas el coeficient de determinació és $R^2 = r_{XY}{}^2 = 0{,}9^2 = 0{,}81$ i per tant, la fiabilitat és bastant elevada.

b) Calculeu el model de regressió lineal que millor aproxima el nombre de treballadors en funció de la facturació.

En aquest cas demana la recta que explica el nombre de treballadors en funció de la facturació. Per tant, ens demana la recta X sobre Y. Per a calcular-la, tan sols cal fer les substitucions corresponents, ja que l'exercici ens dóna tots els estadístics necessaris.

Substituint:

$X - \overline{X} = \frac{S_{XY}}{S_Y^2} \cdot (Y - \overline{Y}) \rightarrow X - 14 = \frac{45}{625} \cdot (Y - 100 \rightarrow$ aïllant la variable X mitjançant matemàtiques elementals, obtenim: X = 0,072 Y + 6,8.

És a dir,

Nombre de treballadors = 0,072 · la facturació + 6,8

c) En funció d'aquest ajust calcula de manera aproximada el nombre de treballadors que s'espera que tinga una empresa que facture 105 milions. És fiable aquesta predicció? Raona la resposta.

Per a fer la predicció únicament cal substituir Y per 105, ja que d'aquesta manera obtindrem l'estimació del nombre de treballadors per a una empresa que tinguera 105 milions de facturació. Així doncs:

Nombre de treballadors = 0,072 · 105 + 6,8 = 14,36

El coeficient de determinació és $R^2 = r_{XY}{}^2 = 0{,}9^2 = 0{,}81$ i per tant, la fiabilitat és bastant elevada.

Exercici 12

Les dues taules següents mostren el grau mitjà de satisfacció dels ocupats segons el treball que realitzen per edat i pel nivell d'estudis l'any 2010. Les dades estan extretes del Ministeri de Treball i Immigració.

NIVELL ESTUDIS	*GRAU DE SATISFACCIÓ*
1	7,05
2	7,09
3	7,21
4	7,23
5	7,50
6	7,55

EDAT	*GRAU DE SATISFACCIÓ*
[16,25)	7,33
[25,30)	7,39
[30,45)	7,37
[45,55)	7,30
[55,65)[10]	7,43

Cal dir que la variable nivell d'estudis ha estat convertida a numèrica discreta per a ser graduable. Així l'equivalència és*: 1 = menys que primaris; 2 = primaris; 3 = secundaris; 4 = batxillerat; 5 = formació professional i 6 = universitaris.* Aquesta conversió s'ha fet a efectes didàctics.

Solució

a) Calculeu el coeficient de relació lineal de totes dues parelles de variables. En quina de les dues convindria calcular la recta de regressió?

Edat i grau de satisfacció

Per a calcular el coeficient de correlació cal fer el mateix que en els exercicis anteriors. No obstant, la variable edat està agrupada i per tant, cal obtenir prèviament les marques de classe.

Anomenem X = edat i Y = grau de satisfacció i calculem els estadístics necessaris per a obtenir r_{XY} .

Així, la taula de la variable X en què apareixen els intervals i les classes, i on es mostren també els productes dels valors de cada variable, així com el sumatori és:

10. En la taula original el darrer interval és 55 anys o més. S'ha tancat l'interval per a fer l'exercici.

X	c_i	Y	$x_i \cdot y_j$
[16,25)	20,5	7,33	150,265
[25,30)	27,5	7,39	203,225
[30,45)	37,5	7,37	276,375
[45,55)	50	7,3	365
[55,65)	60	7,43	445,8
			1440,665

De la taula podem obtenir els estadístics que es necessiten:

$$\overline{X} = 39{,}1 \; S_X = 14{,}413 \text{ i } \overline{Y} = 7{,}364 \; S_Y = 0{,}045$$

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho s'ha de calcular els productes corresponents i fer-ne la suma.

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{20{,}5 \cdot 7{,}33 + ... + 60 \cdot 7{,}43}{5} = \frac{1440{,}665}{5} = 288{,}133$$

i, conseqüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 288{,}133 - 39{,}1 \cdot 7{,}364 = 0{,}2006.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{0{,}2006}{14{,}413 \cdot 0{,}045} = 0{,}309$

Com que el coeficient r_{XY} és tan proper a 0, la relació lineal entre les dues variables és dèbil.

Nivell d'estudis i grau de satisfacció

Anomenem X = nivell d'estudis i Y = grau de satisfacció i calculem els estadístics necessaris per a obtenir r_{XY}:

$$\overline{X} = 3{,}5 \; S_X = 1{,}708 \text{ i } \overline{Y} = 7{,}272 \; S_Y = 0{,}190$$

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho és necessari calcular els productes corresponents i fer-ne la suma.

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{1\cdot 7{,}05 + ... + 6\cdot 7{,}55}{6} = \frac{154{,}58}{6} = 25{,}763$$

i, consegüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 25{,}763 - 3{,}5\cdot 7{,}272 = 0{,}311.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{0{,}311}{1{,}708\cdot 0{,}190} = 0{,}96.$

Com que el coeficient r_{XY} és tan proper a 1, la relació lineal entre les dues variables és forta i té perfecte sentit calcular la recta de regressió.

b) Calculeu la recta de regressió del grau de satisfacció en funció del nivell d'estudis.

Mantenint la notació de l'apartat anterior, X = nivell d'estudis i Y = grau de satisfacció, l'exercici demana la recta de regressió de Y sobre X. Cal, doncs, substituir els estadístics calculats en l'apartat anterior a l'equació de la recta:

$y - \overline{Y} = \dfrac{S_{XY}}{S_X^2}\cdot(x - \overline{X}) \rightarrow y - 7{,}271 = \dfrac{0{,}311}{2{,}917}\cdot(x - 3{,}5) \rightarrow$ aïllant la variable y mitjançant matemàtiques elementals, obtenim: Y = 0,107 X + 6,898.

És a dir,

La satisfació mitjana en el treball = 0,107 · Nivell d'estudis + 6,898

Exercici 13

El grau mitjà de satisfacció dels ocupats segons el treball que realitzen per nivell d'ingressos i per sexe l'any 2010 es mostra en la taula següent. Les dades estan extretes del Ministeri de Treball i Immigració.

NIVELL D'INGRESSOS	*GRAU DE SATISFACCIÓ HOMES*	*GRAU DE SATISFACCIÓ DONES*
[0,600)	6,19	7,253
[600,1000)	6,83	7,234
[1000,1200)	7,28	7,339
[1200,1600)	7,39	7,61
[1600,2100)	7,60	7,768
[2100,3000)	7,82	7,682
[3000,4000)[5]	7,925	7,499

a) Calculeu el coeficient de correlació lineal entre les variables nivell d'ingressos i grau de satisfacció en els homes i entre les variables nivell d'ingressos i grau de satisfacció en les dones. Quines conclusions se'n poden obtenir?

Solució

Si anomenem per X = nivell d'ingressos ; Y = grau de satisfacció mitjà en els homes i Z = grau de satisfacció mitjà en les dones, se'ns demana r_{XY} i r_{XZ}. Per a calcular-los cal fer el mateix que a l'exercici anterior, ja que la variable X està agrupada en intervals i caldrà obtenir les marques de classe. Així:

X	c_i	*Y*	*Z*
[0,600)	**300**	6,19	7,253
[600,1000)	**800**	6,83	7,234
[1000,1200)	**1100**	7,28	7,339
[1200,1600)	**1400**	7,39	7,61
[1600,2100)	**1850**	7,6	7,768
[2100,3000)	**2550**	7,82	7,682
[3000,4000)	**3000**	7,925	7,499

Nivell d'ingressos i grau de satisfacció en els homes

De la taula podem obtenir els estadístics que es necessiten:

$$\overline{X} = 1571{,}429 \; S_X = 889{,}565 \text{ i } \overline{Y} = 7{,}291 S_Y = 0{,}562$$

11. En la taula original el darrer interval és 3000 euros o més. S'ha tancat l'interval per a fer l'exercici.

Cerquem ara $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n}$. Per a fer-ho hem de calcular els productes corresponents i fer-ne la suma.

I per tant:

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{300 \cdot 6{,}19 + ... + 3000 \cdot 7{,}925}{7} = \frac{83451}{7} = 11921{,}571$$

i, consegüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 11921{,}571 - 1571{,}429 \cdot 7{,}291 = 464{,}283.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{464{,}283}{889{,}565 \cdot 0{,}562} = 0{,}93.$

Com que el coeficient r_{XY} és tan proper a 1, la relació lineal entre les dues variables és forta.

Nivell d'ingressos i grau de satisfacció en les doness

En aquest cas els valors dels estadístics són: $\overline{X} = 1571{,}429$; $S_X = 889{,}565$

$\overline{Z} = 7{,}484$; $S_z = 0{,}197$ i $\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{83146{,}9}{7} = 11878{,}129$

i, consegüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 11878{,}129 - 1571{,}429 \cdot 7{,}484 = 117{,}554.$$

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{117{,}554}{889{,}565 \cdot 0{,}197} = 0{,}67$

Com que el coeficient r_{XY} és menor que l'anterior, es pot dir que el grau de satisfacció mitjà en el treball està més relacionat amb el sou per als homes que per a les dones.

b) Calculeu la recta de regressió que explique el grau de satisfacció mitjà en el treball dels homes en funció del nivell d'ingressos.

Mantenint la notació de l'apartat anterior, X = nivell d'ingressos; Y = grau de satisfacció mitjà en els homes, l'exercici demana la recta de regressió de Y sobre X. Cal, doncs, substituir els estadístics calculats en l'apartat anterior a l'equació de la recta:

$y-\overline{Y}=\frac{S_{XY}}{S_X^2}\cdot(x-\overline{X}) \rightarrow y-7{,}291=\frac{464{,}283}{791325{,}889}\cdot(x-1571{,}429)$ → aïllant la variable y mitjançant matemàtiques elementals, obtenim: Y = 0,000587 X + 6,369.

És a dir,

La satisfació mitjana en el treball = 0,000587 · nivell d'ingressos + 6,369

Exercici 14

El nombre total d'expedients de regulació del treball al llarg dels anys 2001-2010, segons les dades estan extretes del Ministeri de Treball i Immigració, són les que es mostren en la taula.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Alacant	139	169	224	268	292	180	164	393	939	679
Castelló	49	59	55	76	88	59	58	291	939	777

Solució

a) Existeix cap tipus de relació lineal entre les variables? És forta aquesta relació? Raoneu les respostes.

Anomenem X = el nombre total d'expedients de regulació a Castelló i per Y = el nombre total d'expedients de regulació a Alacant. Per a saber si existeis cap tipus de relació lineal entre les variables cal estudiar la covariància, i per a saber el grau d'aquesta relació lineal, el coeficient de correlació.

Calculem, doncs, els estadístics necessaris a partir de la taula i de la mateixa manera que fèiem a la unitat 1.

$\overline{X}=344{,}7\ S_X=249{,}694,\ \overline{Y}=245{,}1 S_Y=316{,}02$ i

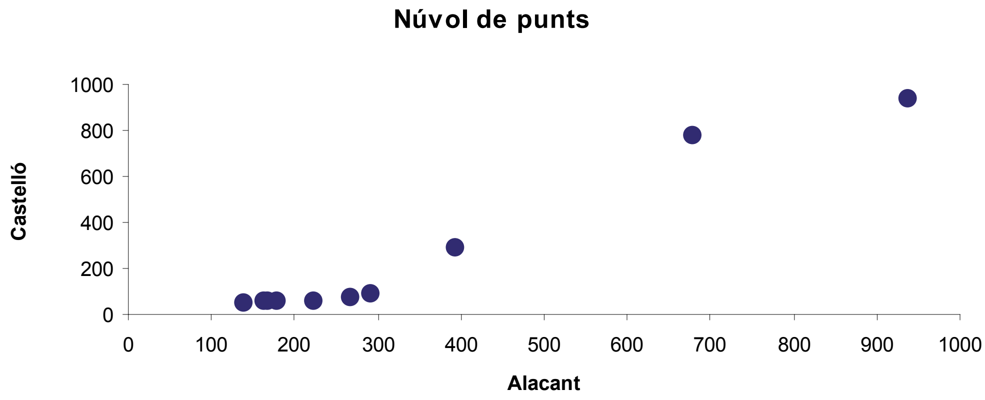
$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j\cdot n_{ij}}{n}=\frac{1618965}{10}=161896{,}5$$

i, conseqüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 161896{,}5 - 344{,}7 \cdot 245{,}1 = 77410{,}53.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{77410{,}53}{249{,}694 \cdot 316{,}02} = 0{,}981.$

Com que el coeficient r_{XY} és molt proper a 1, per tant existeix una correlació lineal molt forta entre totes dues variables, encara que, com es pot observar les dues distribucions marginals tenen molta dispersió. És a dir, malgrat que en les dues variables les mitjanes no siguen representatives de les dades (valors molt alts de la desviació típica) les dades en conjunt sí que «s'amuntonen» al voltant d'una recta. Aquest fet es pot observar en el gràfic de dispersió que apareix tot seguit:



b) Calculeu la recta de regressió lineal que relaciona el nombre d'expedients totals a Alacant en funció dels de Castelló.

Mantenint la notació de l'apartat anterior, de Y sobre X. Cal, doncs, substituir els estadístics calculats en l'apartat anterior a l'equació de la recta:

$y - \overline{Y} = \dfrac{S_{XY}}{S_X^2}\cdot(x - \overline{X}) \rightarrow y - 245{,}1 = \dfrac{77410{,}53}{99868{,}6404}\cdot(x - 249{,}694) \rightarrow$ aïllant la variable *y* mitjançant matemàtiques elementals, obtenim: Y = 0,775 X + 154,713.

És a dir,

expedients d'Alacant = 0,775 · expedients de Castelló + 154,713

Exercici 15

La taula següent mostra el nombre total d'hipoteques signades, així com la taxa d'atur a l'Estat espanyol en el període 2004-2010, segons dades extretes de l'INE.

	Hipoteques	*Taxa d'atur*
2004	1608497	8,1
2005	1798630	9,2
2006	1896515	8,3
2007	1780627	8,6
2008	1283374	13,9
2009	1082587	18,83
2010	961601	20,05

Solució

a) Existeix cap tipus de relació lineal entre les variables? És forta aquesta relació? Raoneu les respostes.

Anomenem Y = nombre d'hipoteques X = taxa d'atur. Per a saber si existeix cap tipus de relació lineal entre les variables cal estudiar la covariància, i per a saber el grau d'aquesta relació lineal, el coeficient de correlació.

Calculem, doncs, els estadístics necessaris a partir de la taula i de la mateixa manera que fèiem a la unitat 1.

$\overline{X} = 12{,}426$; $S_X = 4{,}812$, $\overline{Y} = 1487404{,}429$; $S_y = 347819{,}845$ i

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{118134800{,}26}{7} = 16876400{,}037$$

i, consegüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 16876400{,}037 - 12{,}426 \cdot 1487404{,}429 = \text{-}1606087{,}405.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{\text{-}1606087{,}405}{4{,}812 \cdot 347819{,}845} = 0{,}960.$

Entre aquestes dues variables existeix una relació lineal negativa, ja que la covariància és menor que zero. A més a més, com que el coeficient de correlació està molt prop de –1, la relació lineal és forta.

b) Calculeu la recta de regressió lineal que relaciona el nombre d'hipoteques signades en funció de la taxa d'atur.

Mantenint la notació de l'apartat anterior, de Y sobre X. Cal, doncs, substituir els estadístics calculats l'equació de la recta:

$$y-\overline{Y}=\frac{S_{XY}}{S_X^2}\cdot(x-\overline{X}) \rightarrow y-1487404,429=\frac{-1606087,405}{23,155}\cdot(x-12,426) \rightarrow \text{aïllant}$$

la variable *y* mitjançant matemàtiques elementals, obtenim:

Y = –69362,445 X + 2349302,166, que és l'equació de la recta demanada.

Exercici 16

La taula següent mostra el nombre d'hores extraordinàries totals en milers (remunerades i no remunerades) realitzades en el conjunt de l'Estat espanyol, així com les taxes d'atur des del primer trimestre del 2008 fins al darrer trimeste de l'any 2010. Les dades han estat extretes de l'INE.

Trimestres	*Nombre total d'hores extres*	*Taxa d'atur*
2010TIV	5574,9	20,33
2010TIII	5058,9	19,79
2010TII	6002,7	20,09
2010TI	6154,1	20,05
2009TIV	6493,2	18,83
2009TIII	6069	17,93
2009TII	7042	17,92
2008TIV	8398,4	13,91
2008TIII	8813,2	11,33
2008TII	9794,4	10,44
2008TI	10058,1	9,63

Solució

a) Trobeu, si és pertinent, la recta de regressió que explica el nombre d'hores extres en funció de la taxa d'atur.

Anomenem X = taxa d'atur i Y = nombre d'hores extres. L'exercici demana que calculem la recta de Y sobre X quan siga pertinent, és a dir, quan les dues variables estiguen fortament correlacionades. Per tant, el que cal fer primerament és avaluar el coeficient de correlació entre les dues variables per a decidir si és pertinent o no calcular la recta.

Per a determinar el coeficient de correlació s'han de calcular alguns estadístics, tan de les distribucions marginals com de la conjunta. Així:

$\overline{X} = 16,386$; $S_X = 4,019$; $\overline{Y} = 7223,536$; $S_Y = 1664,837$ i

$$\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} = \frac{1230502,401}{11} = 111863,855$$

i, consegüentment, la covariància és:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 111863,855 – 16,386 \cdot 7223,563 = -6501,448.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{-6501,448}{4,019 \cdot 1664,837} = 0,972.$

Com que el coeficient de correlació està molt pròxim a –1, la relació lineal és forta i té sentit calcular la recta de regressió Y sobre X.

Calculem-la:

$y - \overline{Y} = \dfrac{S_{XY}}{S_X^2}\cdot(x - \overline{X}) \rightarrow y - 7223,536 = \dfrac{-6501,448.}{16,152}\cdot(x - 16,386) \rightarrow$ aïllant la variable y mitjançant matemàtiques elementals, obtenim:

Y = –402,516 X + 13819,163, que és l'equació de la recta demanada.

b) En el primer trimestre de 2009 la taxa d'atur era del 17,36%. Feu una estimació del nombre d'hores extres en aquest trimestre, així com una mesura de la seua fiabilitat.

Per a calcular l'estimació hem de substituir en l'equació de la recta de regressió. Substituint X per 17,36 obtenim:

$$Y = –402,516 \cdot 17,36 + 13819,163 = 6831,485.$$

Per a conèixer la fiabilitat d'aquesta predicció cal determinar el coeficient de determinació[12] R^2, el qual és, en les regressions lineals, el quadrat del coeficient de regressió. Per tant el seu valor és $R^2 = 0,972^2 = 0,945$. Aleshores l'estimació és fiable.

12. Revisar l'exercici 11.

Exercici 17

El nombre total d'expedients de regulació del treball al llarg dels anys 2001-2010 a Catalunya i la Comunitat Valenciana, extrets del Ministeri de Treball i Immigració, són els que es mostren en la taula, a excepció de les dades de l'any 2005 que s'han omés.

ANY	*Catalunya*	*Comunitat Valenciana*
2001	661	465
2002	724	494
2003	608	594
2004	565	619
2006	455	413
2007	470	487
2008	874	1286
2009	3964	3490
2010	3318	2810

Se sap que el nombre d'expedients l'any 2005 a Catalunya va ser de 512. Feu estimació, si convé, del nombre d'expedients a la Comunitat Valenciana així com una mesura de l'ajust.

Solució

Anomenem X = nombre d'expedients a Catalunya i Y = nombre d'expedients a la Comunitat Valenciana. Per a obtenir l'estimació dels expedients a la Comunitat Valenciana a partir de les dades, cal construir la recta de regressió Y sobre X, però perquè aquest càlcul siga profitós, és necessari que les dues variables estiguen fortament correlacionades. Per tant, el que s'ha de fer en primer lloc és calcular el coeficient de correlació.

Per a determinar el coeficient de correlació s'han de calcular alguns estadístics, tant de les distribucions marginals com de la conjunta. Així:

$\overline{X} = 1293{,}2$; $S_X = 1269{,}838$; $\overline{Y} = 1184{,}2$; $S_Y = 1091{,}001$ i la covariància:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 2897179{,}67 - 1293{,}22 \cdot 1184{,}22 = 1365722{,}682.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{1365722{,}682}{1269{,}838 \cdot 1091{,}001} = 0{,}986.$

Com que el coeficient de correlació està molt pròxim a 1, la relació lineal és forta i té sentit calcular la recta de regressió Y sobre X.

Calculem-la:

$y - \overline{Y} = \dfrac{S_{XY}}{S_X^2}\cdot(x - \overline{X}) \rightarrow y - 1184{,}22 = \dfrac{1365722{,}682.}{1269{,}838^2}\cdot(x - 1293{,}22)$ → aïllant la variable y mitjançant matemàtiques elementals, obtenim:

$$Y = 0{,}847\ X + 88{,}907$$

Per a estimar el nombre d'expedients en la Comunitat Valenciana cal substituir X per 512, d'aquesta manera obtenim Y = 0,847 · 512 + 88,907 = 522,571.

Per a calcular la fiabilitat de la predicció hem d'obtenir el coeficient de determinació R^2, el qual és, en les regressions lineals, el quadrat del coeficient de regressió. Per tant, el seu valor és $R^2 = 0{,}986^2 = 0{,}972$. Aleshores, l'estimació és fiable.

Exercici 18

En un museu es vol estudiar la repercussió que tenen les queixes realitzades pels visitants i els ingressos. Per a realitzar-ho, s'observaren les dues variables al llarg de les últimes deu setmanes. Les visites estan expressades en desenes d'assistents.

Queixes	18	26	30	33	38	39	42	44	46	49
Visites	107	105,5	105	104,4	104,3	104	103,7	103,4	103,1	103

Si l'entrada al museu té un cost de 3,6 euros, estimeu els ingressos del museu si en una setmana s'hagueren produït 43 queixes.

Solució

L'exercici demana que estimem els ingressos segons les queixes. Però és evident que els ingressos depenen del nombre de visites, per tant, el que hem d'esbrinar és si existeix cap tipus de relació lineal entre el nombre de queixes i el nombre de visites.

Si efectivament aquesta es produeix, aleshores podrem trobar la recta de regressió entre el nombre de visites i el de queixes i, amb posterioritat, es podran estimar els ingressos.

Així doncs, en primer lloc calcularem el coeficient de correlació entre les dues variables X = nombre de queixes i Y = visites per a saber si estan correlacionades. Fem els càlculs necessaris:

$\overline{X} = 36{,}5$; $S_X = 9{,}211$; $\overline{Y} = 104{,}34$; $S_y = 1{,}166$ i la covariància:

$$S_{XY} = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{x_i y_j \cdot n_{ij}}{n} - \overline{X}\cdot\overline{Y} = 3797{,}82 - 36{,}5 \cdot 104{,}34 = -10{,}59.$$

I per tant, $r_{XY} = \dfrac{S_{XY}}{S_X \cdot S_Y} = \dfrac{-10{,}59}{9{,}211 \cdot 1{,}166} = -0{,}986.$

Com que el coeficient de correlació està molt pròxim a –1, la relació lineal és forta i té sentit calcular la recta de regressió Y sobre X.

Calculem-la:

$y - \overline{Y} = \dfrac{S_{XY}}{S_X^2}\cdot(x - \overline{X}) \rightarrow y - 104{,}34 = \dfrac{-10{,}59.}{9{,}211^2}\cdot(x - 36{,}5) \rightarrow$ aïllant la variable y mitjançant matemàtiques elementals, obtenim:

$$Y = 0{,}125\ X + 108{,}866.$$

Per a estimar el nombre de visites quan es produeixen 43 queixes cal únicament substituir X per 43, d'aquesta manera obtenim $Y = 0{,}125 \cdot 43 + 108{,}866 = 114{,}241$.

Per a calcular la fiabilitat de la predicció hem d'obtenir el coeficient de determinació R^2, el qual és, en les regressions lineals, el quadrat del coeficient de regressió. Per tant, el seu valor és $R^2 = 0{,}986^2 = 0{,}972$. L'estimació, doncs, és fiable.

Finalment, els ingressos estimats són 3,6 euros multiplicat pel nombre de visites estimades. És a dir, $3{,}6 \cdot 1088{,}66 = 3919{,}176$ euros.

Exercici 19

La taula següent mostra el nombre de persones ocupades distribuïdes atenent el sou net de l'activitat principal que desenvolupen (en centenes d'euro) i el nivell d'estudis que tenien l'any 2010, segons dades recollides del Ministeri de Treball i d'Immigració. Cal dir però, que la variable nivell d'estudis ha estat convertida a numèrica discreta per ser graduable. Així l'equivalència és*: 1 = menys que primaris; 2 = primaris; 3 = secundaris; 4 = batxillerat; 5 = formació professional i 6 = universitaris.* Aquesta conversió s'ha fet a efectes didàctics.

	SOU						
Nivell d'estudis	***[0, 6)***	***[6,10)***	***[10,12)***	***[12,16)***	***[16,21)***	***[21,30)***	***[30, 40)***
1	8,75	21,67	*15,00*	7,93	5,65	*0,74*	1,56
2	293,18	790,92	*601,61*	472,64	92,82	*56,30*	3,77
3	538,08	1551,52	*1226,20*	1098,34	340,72	*74,78*	13,56
4	323,39	670,29	*607,31*	709,62	313,35	*142,00*	53,20
5	303,28	801,87	843,80	982,42	444,90	183,90	50,73
6	164,47	439,26	619,51	1155,75	1230,07	919,40	282,14

Solució

a) Estan relacionats linealment el sou i el nivell d'estudis?

Anomenem X = nivell d'estudis i Y = sou. Hem de calcular el coeficient de correlació entre totes dues variables per saber si les dues variables estan linealment relacionades i amb quin grau.

Per a fer aquests càlculs cal tenir present que la variable Y està agrupada en intervals i necessitem les marques de classe, i que cada dada bivariant té una freqüència absoluta superior a 1. Aquest darrer fet pot complicar els càlculs, per això és convenient seguir un criteri a l'hora de fer-los. Nosaltres ho farem en la taula de doble entrada:

X	[0, 6)	[6,10)	[10,12)	[12,16)	[16,21)	[21,30)	[30, 40)	
Marques-->	**3**	**8**	**11**	**14**	**18,5**	**25,5**	**35**	***n.j***
1	8,75	21,67	*15*	7,93	5,65	*0,74*	1,56	***61,3***
2	293,18	790,92	*601,61*	472,64	92,82	*56,3*	3,77	***2311,24***
3	538,08	1551,52	*1226,2*	1098,34	340,72	*74,78*	13,56	***4843,2***
4	323,39	670,29	*607,31*	709,62	313,35	*142*	53,2	***2819,16***
5	303,28	801,87	843,8	982,42	444,9	183,9	50,73	***3610,9***
6	164,47	439,26	619,51	1155,75	1230,07	919,4	282,14	***4810,6***
$n_{i.}$	***1631,15***	***4275,53***	***3913,43***	***4426,7***	***2427,51***	***1377,12***	***404,96***	**18456,4**
$\square\, x_i \cdot n_{i1}$	***6006,13***	***15584,14***	***15262,12***	***18933,31***	***12071,77***	***7341,58***	***2209,07***	
$y_1 \square\, x_i \cdot n_{i1}$	***18018,4***	***124673,12***	***167883,32***	***265066,3***	***223327,75***	***187210,29***	***77317,5***	**1063496,7**

De la manera natural es calculen els estadístics de les distribucions marginals: $\overline{X} = 4{,}194$; $S_X = 1{,}412$; $\overline{Y} = 12{,}913$; $S_Y = 6{,}478$. Per a calcular la covariància cal fer els productes de cada valor pel seu corresponent i després fer-ne la suma. Així, el procediment que emprem és el següent: la fila amb $\sum x_i \cdot n_{i1}$ s'obté multiplicant cada fila per la seua freqüència conjunta, per exemple, 6006,13 = 1 · 8,75 + 2 · 293,18 + ...+ 6 · 164,47. La fila $y_1 \sum x_i \cdot n_{i1}$ sóbté multiplicant cada resultat de $\sum x_i \cdot n_{i1}$ pel valor de la primer columna,; per exemple, 18018,4 = 6006,13 · 3. Així doncs, $\sum_{i=1}^{h} \sum_{j=1}^{k} \frac{x_i y_j \cdot n_{ij}}{n} = \frac{1063496{,}7}{18456{,}4} = 57{,}622$. Com a conseqüència sóbté que la covariància és $S_{XY} = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{x_i y_j \cdot n_{ij}}{n} - \overline{X} \cdot \overline{Y} = 57{,}622 - 4{,}194 \cdot 12{,}913 = 3{,}465$.

I per tant, $r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{3{,}465}{1{,}412 \cdot 6{,}478} = 0{,}38$, el qual és molt baix, i aleshores la relació és molt dèbil i no tindria gaire sentit fer la recta de regressió.

b) Feu una estimació del sou que cobraria una persona ocupada que tinguera un nivell d'estudis equivalent a 4,5 així com la seua fiabilitat.

Com s'ha comentat en l'apartat anterior, el coeficient de correlació és molt baix i per tant manca de sentit fer la recta de regressió. No obstant, la calcularem per a fins didàctics.

Hem de calcular la recta Y sobre X. Llavors:

$y - \overline{Y} = \frac{S_{XY}}{S_X^2} \cdot (x - \overline{X}) \rightarrow y - 12{,}913 = \frac{3{,}465.}{1{,}412^2} \cdot (x - 4{,}194) \rightarrow$ aïllant la variable y mitjançant matemàtiques elementals, obtenim:

$$Y = 1{,}738\ X + 5{,}622.$$

Per a estimar el sou d'una persona amb un nivell d'estudis de 4,5: Y = 1,738 4,5 + 5,622 = 13,442 milers d'euros.

Per a calcular la fiabilitat de la predicció hem d'obtenir el coeficient de determinació R^2, el qual és, en les regressions lineals, el quadrat del coeficient de regressió. Per tant, el seu valor és $R^2 = 0{,}38^2 = 0{,}144$. L'estimació, doncs, és molt poc fiable.

UNITAT 3

Nombres índexs

Introducció teòrica

Com a elements introductoris d'aquest capítol, és convenient recordar definicions de conceptes que necessitarem per a assolir els objectius a treballar en aquesta unitat (referències bibliogràfiques 7, 11 i 18).

Índex simple

Els índexs, calculats a partir d'una sèrie de dades d'una magnitud i en un període *t* que denotarem per x_{it}, ens permetran avaluar, en termes relatius o percentatges, l'evolució de les dades de la sèrie per períodes. Així:

$$I_{it-1}^{t} = \frac{x_{it}}{x_{it-1}}$$

Índex complex

Parlarem d'índex complex, quan volem estudiar l'evolució d'una magnitud complexa, perquè ens interessa aglutinar en una de sola, la multiplicitat de diverses magnituds simples. En un exemple per a analitzar la producció de cereals en certa comunitat (magnitud complexa) necessitem les dades de panís, blat, civada, etc. (que serien les magnituds simples).

Dintre dels índexs complexos, nosaltres treballarem amb índexs ponderats. Veurem als exercicis les fórmules dels índexs de preus i quantitats de Laspeyres i Paasche, amb la intenció d'analitzar les raons que justifiquen la utilització de l'índex de Laspeyres de preus en el càlcul oficial de l'IPC a Europa. Així:

Índex de preus per Laspeyres

Per a calcular aquest índex, començarem per conèixer i deduir la fórmula a emprar. Hem reduït el seu càlcul a tres articles però no oblidem que aquest càlcul s'estén a la totalitat d'articles representatius del consum de les famílies en un país (Veure ECPF).

$$L_{preus}^{t} = \frac{\sum_{i} \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_{i} p_{i0} \cdot q_{i0}} = \frac{\sum_{i} p_{it} \cdot q_{i0}}{\sum_{i} p_{i0} \cdot q_{i0}}$$

on *t* és l'any actual i *0* serà l'any que prendrem com a referència en la comparació. Si es tracta d'índexs encadenats podríem dir-ne anys *t-1* i *t*:

p_{it} serà el preu de l'article i a l'any t
p_{i0} serà el preu de l'article i a l'any 0
q_{it} serà la quantitat de l'article i a l'any t
q_{i0} serà la quantitat de l'article i a l'any 0

és una mitjana ponderada on el «pes» de cada article $p_{i0} \cdot q_{i0}$ és el valor de l'article en la «cistella de la compra» de l'any de referència i romandrà constant al llarg del període mentre no se'n canvie la base. Un inconvenient d'aquest mètode és que si la importància dels articles en els hàbits de consum canvia molt, aquests coeficients queden desfasats.

Índex de quantitats per Laspeyres

En aquest cas s'estudia l'evolució de les quantitats demandades i per a la ponderació s'utilitzen els mateixos coeficients de la fórmula anterior $p_{i0} \cdot q_{i0}$:

$$L_{quantitats}{}_0^t = \frac{\sum_i \frac{q_{it}}{q_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_i p_{i0} \cdot q_{i0}} = \frac{\sum_i q_{it} \cdot p_{i0}}{\sum_i p_{i0} \cdot q_{i0}}$$

Índex de preus per Paasche

$$P_{preus}{}_0^t = \frac{\sum_i \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{it}}{\sum_i p_{i0} \cdot q_{it}} = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i p_{i0} \cdot q_{it}}$$

és una mitjana ponderada on el «pes» de cada article $p_{i0} \cdot q_{it}$ intenta millorar la proposta de Laspeyres, evitant en certa manera el desfassament, ja que arreplega la importància de l'article en considerar la quantitat en el període a comparar. Presenta altres inconvenients.

Índex de quantitats per Paasche

$$P_{quantitats}{}_0^t = \frac{\sum_i \frac{q_{it}}{q_{i0}} \cdot q_{i0} \cdot p_{it}}{\sum_i q_{i0} \cdot p_{it}} = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i q_{i0} \cdot p_{it}}$$

Aquesta proposta, com que analitza l'evolució de les quantitats, considera com a coeficient $p_{i0} \cdot q_{it}$ que indica el «pes» de cada article i, el preu de l'any t per a actualitzar la importància de l'article.

Índexs encadenats

Per a calcular l'increment d'una sèrie en un període més llarg, podem utilitzar els índexs prèviament coneguts dels períodes que internament conformen el període total. Així:

$$I_t^m = I_t^{t+1} \cdot I_{t+1}^{t+2} \cdot I_{t+2}^{t+3} \cdot \ldots \cdot I_{m-2}^{m-1} \cdot I_{m-1}^m$$

Increment mitjà

En cert període llarg que inclou diversos períodes menors internament, podem conèixer els increments del períodes menors, un a un, consecutivament i observar entre ells acusades diferències en signes (augments i disminucions) o en magnitud. Per a calcular l'increment mitjà de tots ells podem fer l'arrel del producte d'aquests índexs, tenint en compte que l'índex de l'arrel coincideix amb el nombre d'índexs considerats. Així:

$$I_{t\,mitjà}^{m} = \sqrt[m-t]{I_t^{t+1} \cdot I_{t+1}^{t+2} \cdot I_{t+2}^{t+3} \cdot \ldots \cdot I_{m-2}^{m-1} \cdot I_{m-1}^m}$$

Canvi de base a l'IPC

Per raons que caldrà aprofundir en la teoria, en certs moments calia fer un canvi en l'any de referència del càlcul oficial de l'IPC i es començava a obtenir la nova sèrie d'aquest índex, començat de nou amb el valor 100. Direm que es feia un «canvi de base».

Sovint, com podrem veure en els exercicis proposats, cal utilitzar en un mateix càlcul el valor de l'IPC d'anys que corresponen a períodes de bases diferents i necessitarem treballar amb tots els valors de l'IPC referits a una mateixa base. Aquestes dades les podràs trobar fàcilment, a la pàgina web de l'INE però als exercicis podrem veure com es pot trobar l'enllaç tècnic que permet unificar la base de les dues sèries.

Per a obtenir el valor a l'IPC de l'any «i» en base B hem multiplicat l'IPC de l'any «i» en base A per la fracció $\dfrac{IPC_B^B}{IPC_A^B}$ el valor de la qual es el que denominarem «enllaç tècnic».

Seria interessant aprofundir en els mecanismes del càlcul de l'IPC per l'INE i els canvis metodològics introduïts en els darrers anys a partir del 2000 i el procés d'harmonització amb Europa.

http://www.ine.es/daco/daco43/metoipc06.pdf

Actualització d'un valor

Utilitzant els valors de l'IPC, podem conèixer el valor actualitzat d'una renda, d'un lloguer, d'un sou, d'un bé, etc. Tan sols hem considerat que aquesta operació la podrem fer sempre que els períodes inicials i finals a considerar estiguen ambdós abans o després del gener del 2002. En altre cas, caldrà recórrer a un índex LAU que es pot trobar a l'INE:

$$Valor\ \ actualitzat = Valor\ \ inicial\ \times \frac{IPC\ \ mes\ \ final}{IPC\ \ mes\ \ inicial}$$

Pèrdua o guany de poder adquisitiu

Es pot dir que nosaltres guanyem poder adquisitiu (capacitat de compra de béns de consum) si el salari que percebem aquest any està per damunt del que percebríem si el nostre salari haguera estat incrementat en el mateix percentatge que augmenten els preus d'aquests béns. Podríem raonar de la mateixa manera per definir la pèrdua de poder adquisitiu quan el nostre salari queda per sota del que tindríem si l'hagueren incrementat amb el mateix percentatge que els preus.

L'increment d'aquests preus està reflectit a l'IPC que publica l'INE cada mes. Nosaltres agafarem la mitjana anual general d'aquest índex que podrem trobar fàcilment al web d'aquest organisme.

Ara bé, com fem una anàlisi en termes relatius i donem el resultat en percentatges, vegem en la següent expressió, com el salari concret del que partim, no és necessari en l'estudi de l'evolució del poder adquisitiu:

Guany o pèrdua de poder adquisitiu

$$= \boldsymbol{\Delta}_{poder\ adquisitiu} = \frac{\textbf{\textit{salari nou}}}{\textbf{\textit{salari actualitzat segons IPC}}} = \frac{\textbf{\textit{salari anterior}} \cdot (1 + \boldsymbol{\Delta}_{salarial})}{\textbf{\textit{salari anterior}} \cdot (1 + \boldsymbol{\Delta}_{IPC})} =$$

$$= \frac{(1 + \boldsymbol{\Delta}_{salarial})}{(1 + \boldsymbol{\Delta}_{IPC})}$$

La pèrdua o guany del poder adquisitiu, doncs, es calcula a partir de la comparació dels increments anuals del salari (Δ_{salari}) i de l'IPC (Δ_{IPC}) paral·lelament.

A tal fi, començarem per calcular els índexs «encadenats» dels salaris, que ens permetran esbrinar els increments salarials anuals i a partir dels valors de l'IPC publicats a l'INE podrem fer-hi el mateix.

Deflactar una sèrie o passar-la a termes reals

Per a estudiar l'anàlisi d'una magnitud econòmica en *termes reals*, cal transformar els valors originals en *termes corrents* mitjançant els IPC que convinguen, per a convertir tots els valors de la sèrie en els equivalents referits a un mateix any que denominarem *any de referència*. Aquesta operació s'anomena *deflactació* de la sèrie.

Amb aquesta operació, li hem «eliminat» a la sèrie original, l'efecte de la inflació i podrem analitzar «en termes reals» la seua evolució com a tal magnitud, excepte les influències dels esdeveniments de l'economia general que es reflecteixen en les variacions de l'índex de preus.

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

- 3*a*) Saber calcular els nombres índexs simples d'una sèrie de valors per a estudiar l'evolució d'una magnitud al llarg del temps.
- 3*b*) Interpretar el valor de l'índex per a conèixer l'increment percentual de la magnitud en el període indicat i viceversa.
- 3*c*) Saber calcular índexs amb la mateixa base de referència i també índexs encadenats.
- 3*d*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com els seus increments percentuals per períodes.
- 3*e*) Conèixer i calcular l'enllaç per a canviar de base els índexs.
- 3*f*) Conèixer les fórmules de Laspeyres i Paasche com a índexs complexos.
- 3*g*) Actualitzar el valor d'un bé, utilitzant els valors de l'IPC.
- 3*h*) Deflactar una sèrie utilitzant l'IPC.
- 3*i*) Conèixer el concepte de termes monetaris nominals (moneda corrent) i termes reals (moneda constant) en una sèrie econòmica per a avaluar la seua evolució.
- 3*j*) Fer previsions dels valors d'una sèrie per a dades immediates.
- 3*k*) Saber calcular les variacions del poder adquisitiu d'un salari, en funció de les variacions del salari i de l'IPC.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

Objectiu **Exercici**	***3a***	***3b***	***3c***	***3d***	***3e***	***3f***	***3g***	***3h***	***3i***	***3j***	***3k***
1	**x**	**x**	**x**	**x**						x	
2	x	x	x	x						x	
3	x	x	x	x						x	
4					x						
5						x					
6							x				
7				x				x	x	x	
8											x
9											x
10	x	x	x	x	x			x	x	x	

Enunciats

3*a*) Saber calcular els nombres índexs simples d'una sèrie de valors per a estudiar l'evolució d'una magnitud al llarg del temps.
3*b*) Interpretar el valor de l'índex per a conèixer l'increment percentual de la magnitud en el període indicat i viceversa.
3*c*) Saber calcular índex amb la mateixa base de referència i també índex encadenats.
3*d*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com si coneixem els seus increments percentuals per períodes.
3*j*) Fer previsions dels valors d'una sèrie per a dades immediates.

Exercici 1

A continuació presentem el volum total d'alumnes matriculats a la Universitat Jaume I els darreres anys.

	Nombre total alumnes matriculats
Curs 2005/2006	12676
Curs 2006/2007	12928
Curs 2007/2008	13159
Curs 2008/2009	13210
Curs 2009/2010	13904
Curs 2010/2011	14702

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2005 (farà referència al curs 2005/2006). Interpreteu-ne el resultat.
b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà anual d'aquest període, a partir de les quantitats originals i dels índexs encadenats.
d) Feu previsions per a la matrícula del curs 2011/2012 i 2012/2013 si considerem que no hi haurà canvis significatius en el seu comportament.

3*a*) Saber calcular els nombres índexs simples d'una sèrie de valors per a estudiar l'evolució d'una magnitud al llarg del temps.

3*b*) Interpretar el valor de l'índex per a conèixer l'increment percentual de la magnitud en el període indicat i viceversa.

3*c*) Saber calcular índexs amb la mateixa base de referència i també índexs encadenats.

3*d*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com els seus increments percentuals per períodes.

3*j*) Fer previsions dels valors d'una sèrie per a dades immediates.

Exercici 2

En la següent taula es mostren les dades de l'INI que fan referència al total de visitants en els parcs nacionals d'Espanya, en els anys que fem referència.

Natura i biodiversitat	
Zones protegides	
Nombre de visitants per parcs nacionals i període	
Unitats: nombre de persones	
	Total
2000	10252799
2001	10002517
2002	9661493
2003	10296382
2004	11134880
2005	10743480
2006	10979470
2007	10864738
2008	10222818
2009	9952606
Font:Ministeri de Medi Ambient i Medi Rural i Marí. Xarxa de Parcs Naturals Copyright INE 2011	

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2000 i interpreteu-ne el resultat.
b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà anual d'aquest període, a partir de les quantitats originals i dels índexs encadenats.
d) Feu previsions del nombre de visitants dels parcs considerats per als anys 2010, 2011 i 2012, si considerem que no hi haurà canvis significatius en el comportament de l'afluència.

Font: INE

3*a*) Saber calcular els nombres índexs simples d'una sèrie de valors per a estudiar l'evolució d'una magnitud al llarg del temps.
3*b*) Interpretar el valor de l'índex per a conèixer l'increment percentual de la magnitud en el període indicat i viceversa.
3*c*) Saber calcular índexs amb la mateixa base de referència i també índexs encadenats.
3*d*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com els seus increments percentuals per períodes.
3*j*) Fer previsions dels valors d'una sèrie per a dades immediates.

Exercici 3

A continuació presentem les variacions percentuals del volum de vendes de certa superfície comercial, els darrers anys.

Any	**Variacions del volum de vendes (%)**
2006	–3,13
2007	–2,15
2008	+2,12
2009	+3,15
2010	+4,12
2011	+4,31

a) Calculeu els índexs de les vendes de cada any, prenent com a referència l'any 2005 i els índexs encadenats.
b) Calculeu la variació o increment mitjà anual i total de les vendes en aquest període.
c) Estimeu les vendes dels dos anys següents si suposem que no hi ha canvis significatius en el comportament de les vendes en aquests anys.

3*e*) Conèixer i calcular l'enllaç per a canviar de base els índexs.

Exercici 4

A continuació presentem els valors de l'índex de preus de consum, IPC, que podem consultar en la pàgina de l'INE, que fa referència a les dades en base al 2001 i 2006. Per raons que caldrà estudiar en la teoria, en certs moments s'ha de fer un canvi en l'any de referència i es comença a obtenir la nova sèrie de l'IPC, iniciada de nou amb el valor 100. Direm, aleshores, que hi ha hagut un «canvi de base».

Sovint, com podreu veure en exercicis posteriors, cal utilitzar en un mateix càlcul el valor de l'IPC d'anys que corresponen a períodes de bases diferents i necessitarem treballar amb tots els valors de l'IPC referits a una mateixa base. Aquestes dades les podreu trobar fàcilment en la pàgina web de l'INE, però en aquest exercici anem a veure com es calculen els valors de les caselles que estan ombrejades amb gris.

En primer lloc, presentem la taula dels valors de l'IPC des de l'any 2002 al 2006 en base 2001.

Índex de preus de consum	
Mitjanes anuals. Base 2001	
Nacional per general i grups COICOP	
Unitats: índexs i tasses	
	General
	Mitjana anual
2006	117,624
2005	113,63
2004	109,927
2003	106,684
2002	103,538

I a continuació, les dades dels valors de l'IPC des de l'any 2006 al 2010 en base 2006, encara que estan afegits els valors de les caselles grisses que corresponen als valors obtinguts a «posteriori» per a facilitar els treballs de càlcul referits a períodes de diferents bases.

Índex de preus de consum	
Mitjanes anuals. Base 2006	
Índexs nacionals: general i de grups COICOP	
Unitats: Base 2006 = 100	
	General
	Mitjana anual
2010	108,588
2009	106,668
2008	106,976
2007	102,787
2006	100
2005	96,604
2004	93,456
2003	90,699
2002	88,024

Expliqueu com s'han obtingut les dades de les caselles ombrejades en gris, esbrinant el valor de l'enllaç.

Font: INE

3*f*) Conèixer les fórmules de Laspeyres i Paasche com a índexs complexos.

Exercici 5

Calculeu els índexs de preus i quantitats dels articles A, B i C mitjançant la fórmula de Laspeyres i Paasche, dels anys 2008, 2009 i 2010 en funció de l'any 2008, utilitzant les dades de les taules següents on estan indicades les quantitats q_i i preus p_i que cal conèixer.

	2008		2009		20010	
	Preu p_i	Quantitat q_i	Preu p_i	Quantitat q_i	Preu p_i	Quantitat q_i
Art. A	12	100	14	112	15	115
Art. B	10	50	8	65	7	72
Art. C	5	20	10	10	15	5

3*g*) Actualitzar el valor d'un bé, utilitzant els valors de l'IPC.

Exercici 6

Suposem que vam comprar un habitatge per 16125000 de pesetes el mes de desembre de 1998 i l'hem venut el mes de desembre de 2006 per un valor de 240000 euros. Esbrina el percentatge de beneficis o pèrdues que hem tingut en l'operació.

Nota: Per a fer les operacions consultarem els valors de l'IPC que necessitem a la pàgina web de l'INE. www.ine.es (seria interessant calcular aquest increment amb l'IPC general i amb l'IPC del grup de l'habitatge).

El canvi de moneda que considerarem es 1 € = 166,386 PTA

3*h*) Deflactar una sèrie utilitzant l'IPC.
3*i*) Conèixer el concepte de termes monetaris nominals (moneda corrent) i termes reals (moneda constant) en una sèrie econòmica per a avaluar la seua evolució.
3*j*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com els seus increments percentuals per períodes.
3*k*) Fer previsions dels valors d'una sèrie per a dades immediates als valors coneguts.

Exercici 7

A la taula següent mostrem les dades dels impostos municipals de cert habitatge els darrers anys.

Any	Import impost municipal (termes nominals)
2006	503,24
2007	515,65
2008	536,73
2009	578,84
2010	584,42

Per a analitzar la seua evolució:

a) Deflacteu la sèrie, convertint-la a monedes constants de l'any 2006.
b) Calculeu els índexs que ens permetran estudiar la seua evolució any per any, en termes reals o monedes constants de l'any 2006. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà en el període en termes reals.
d) Suposant que els impostos seguisquen aquest comportament, esbrineu el valor, en termes nominals o monedes corrents per als anys 2011, 2012 i 2013.

Nota: per a resoldre aquest exercici, utilitzarem els valors de la mitjana anual de l'IPC general que necessitem, obtenint-los de la pàgina web de l'INE: www.ine.es

3*k*) Saber calcular les variacions del poder adquisitiu d'un salari, en funció de les variacions del salari i de l'IPC.

Exercici 8

En la taula següent indiquem el valor de la nòmina mensual d'un treballador els darrers anys.

Estudieu la pèrdua o guany del seu poder adquisitiu per a cada any i de tot el període global, considerant els valors de la mitjana anual de l'IPC general que pots trobar en la pàgina de l'INE.

Any	Nòmina mensual (€)
2007	2034,75
2008	2062,13
2009	2218,61
2010	2253,67
2011	2181,75

3*k*) Saber calcular les variacions del poder adquisitiu d'un salari, en funció de les seues variacions i de l'IPC.

Exercici 9

En les taules següents es presenten els valors de l'IPC i l'increment salarial d'un treballador dels anys que indiquem en certa comunitat.

Anys	IPC
2008	115,1
2009	119,2
2010	121,6
2011	123,8

Anys	Increment salarial anual (%) Anual IPC
2008	
2009	1,8
2010	2,7
2011	1,7

a) Calculeu l'increment mitjà i total del salari en el període 2008-2011.
b) Calculeu l'increment anual, mitjà i total de l'IPC en el període 2008-2011.
c) Si suposem que les condicions econòmiques de la comunitat no varien, realitza una previsió del valor de l'IPC per a l'any 2013.
d) Estudieu per a cada any i per al període total la pèrdua o guany del poder adquisitiu i realitzeu una interpretació de les dades obtingudes.

3*a*) Saber calcular els nombres índexs simples d'una sèrie de valors per a estudiar l'evolució d'una magnitud al llarg del temps.
3*b*) Interpretar el valor de l'índex per a conèixer l'increment percentual de la magnitud en el període indicat i viceversa.
3*c*) Saber calcular índexs amb la mateixa base de referència i també índexs encadenats.
3*d*) Calcular l'increment total i mitjà d'una sèrie en cert període, així com els índexs corresponents, tant si coneixem els termes d'una sèrie com els seus increments percentuals per períodes.
3*e*) Conèixer i calcular l'enllaç per a canviar de base els índexs.
3*h*) Deflactar una sèrie utilitzant l'IPC.
3*i*) Conèixer el concepte de termes monetaris nominals (moneda corrent) i termes reals (moneda constant) en una sèrie econòmica per a avaluar la seua evolució.
3*j*) Fer previsions dels valors d'una sèrie per a dades immediates.

Exercici 10

Per a fer un estudi de l'evolució del preu de cert model d'ordinador en termes reals, disposem de les dades que presentem en la taula següent:

a) Calculeu l'increment anual, mitjà i total del preu de l'ordinador en termes reals.
b) Si seguim aquesta evolució, estimeu el preu que podria tenir l'ordinador l'any 2008.

c)

	2000	2001	2002	2003	2004
IPC base 1992	131	135	139		
IPC base 2002			103	106	109
	1300	1275	1250	1100	950

NOTA: Hem de recórrer a períodes i valors molt antics o imaginats per a treballar l'objectiu del canvi de base de l'IPC, a causa de la nova metodologia del càlcul de l'IPC per l'INE aquesta circumstància s'ha superat, pero és important que l'alumnat conega aquest contingut per a advertir la necessitat de no treballar en sèries d'IPC no adients en un mateix exercici.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de realitzar els exercicis i problemes. És convenient no fer-ne un abús excessiu, és a dir, abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10-15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2, i en darrer terme, la solució.

Ajudes Tipus 1

Exercici 1

En l'apartat *a*) s'han de calcular els índexs en base 2005, comparant de manera percentual cada valor de la llista amb 12676.

En l'apartat *b*) cal fer el mateix que en l'apartat anterior però comparant cada valor de la llista amb l'anterior, utilitzant i interpretant els índexs corresponents.

En l'apartat *c*) cal comparar el primer i darrer valor de la llista per a l'increment total i calcular després l'arrel corresponent per a determinar l'increment mitjà anual.

Cal decidir l'índex d'aquesta arrel, en funció del nombre d'anys que hi considerem en el període.

d) Per a fer previsions de la matrícula per als cursos propers, utilitzarem l'increment mitjà anual que hem obtingut en l'apartat anterior.

Exercici 2

Tots els apartats d'aquest exercici són iguals que els de l'exercici anterior excepte en un aspecte: al primer exercici, els valors de la magnitud sempre creixien, mentre que en aquest segon exercici alguns valors augmenten i d'altres disminueixen al llarg del període considerat.

Exercici 3

En aquest exercici, a diferència dels dos anteriors, no coneixem els valors de la magnitud que volem estudiar. A l'enunciat ens donen els percentatges de creixement o decreixement directament. Cal que convertim aquesta informació directament en índex. Per exemple, pots emplenar la següent taula per a fer l'apartat *a*) i resoldre la resta d'apartats com ho has aprés als exercicis 1 i 2.

Any	Variacions del volum de vendes (%)	Índexs Encadenats
2006	-3,13	$I_{2005}^{2006} = 0{,}9687$
2007	*-2,15*	$I_{2006}^{2007} =$
2008	+*2,12*	$I_{2007}^{2008} = 1{,}0212$
2009	+*3,15*	$I_{2008}^{2009} =$
2010	+*4,12*	$I_{2009}^{2010} =$
2011	+*4,31*	$I_{2010}^{2011} =$

És interessant aprendre a calcular els índexs en base 2005 a partir dels índexs encadenats de la taula anterior. Per exemple:

$$I_{2005}^{2007} = I_{2005}^{2006} \cdot I_{2006}^{2007} = 0{,}9687 \cdot 0{,}9785 = 0{,}9479$$

i així successivament...

$$I_{2005}^{2008} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} = \ldots$$

$$I_{2005}^{2009} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} = \ldots$$

$$I_{2005}^{2010} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} = \ldots$$

$$I_{2005}^{2011} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} \cdot I_{2010}^{2011} = \ldots$$

En l'apartat *b*) ens demanen l'increment total que ja hem trobat en la darrera línia i determinar l'increment mitjà anual mitjançant el càlcul de l'arrel sisena de l'índex que acabem de nomenar.

En l'apartat *c*) cal fer una reflexió. Expliqueu per què no és possible fer la previsió que ens demanen.

Exercici 4

En aquest exercici ens demanen un estudi de com es fa *un canvi de base* en la seqüència de valors de l'IPC. Recordeu que aquest càlcul és necessari fer-lo sempre que necessitem utilitzar en una mateixa operació valors de l'IPC que corresponen a anys que s'han calculat amb diferents bases. També parlarem de l'*enllaç tècnic* que permet fer directament aquesta transformació.

Cal comprendre que tan sols es planteja una proporcionalitat (regla de tres) que es manté entre la seqüència de valors de la sèrie d'una base i la sèrie obtinguda amb la base nova. També s'ha de considerar que l'any en què es fa el canvi de base, es calcula l'IPC amb la base vella i la nova, i es planteja l'equivalència.

Vegeu el plantejament del primer any que cal resoldre:

$$IPC_{2001}^{2002} = 103{,}538 \rightarrow IPC_{2006}^{2002} = \boldsymbol{x}$$
$$IPC_{2001}^{2006} = \boldsymbol{117{,}624} \rightarrow IPC_{2006}^{2006} = \boldsymbol{100}$$

Les dades marcades en roig seran els termes necessaris per a definir l'enllaç tècnic.

Exercici 5

Per a resoldre aquest exercici és necessari conèixer les fórmules a emprar, les quals podeu consultar en un manual de teoria sobre els índexs de preus i quantitats.

Índex de preus per Laspeyres

$$L_{preus}{}_0^t = \frac{\sum_i p_{it} \cdot q_{i0}}{\sum_i p_{i0} \cdot q_{i0}}$$

Índex de quantitats per Laspeyres

$$L_{quantitats}{}_0^t = \frac{\sum_i q_{it} \cdot p_{i0}}{\sum_i p_{i0} \cdot q_{i0}}$$

Índex de preus per Paasche

$$P_{preus}{}_0^t = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i p_{i0} \cdot q_{it}}$$

Índex de quantitats per Paasche

$$P_{quantitats}{}_0^t = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i q_{i0} \cdot p_{it}}$$

on *t* és l'any actual i *0* serà l'any que prendrem com a referència en la comparació. Si es tracta d'índexs encadenats podríem anomenar-los anys *t-1* i *t*.

p_{it} serà el preu de l'article i *a* l'any *t*
p_{i0} serà el preu de l'article i *a* l'any 0
q_{it} serà la quantitat de l'article i *a* l'any *t*
q_{i0} serà la quantitat de l'article i *a* l'any 0

Exercici 6

En aquest exercici, en primer lloc, cal actualitzar la moneda. Nosaltres ho hem resolt aplicant el canvi i convertint la quantitat de compra a euros.

Per a actualitzar el valor, s'ha de conèixer el valor de l'IPC del moment de compra i venda en la mateixa base i operar convenientment.

Exercici 7

En aquest exercici, aprendrem a estudiar l'anàlisi d'una magnitud econòmica en termes reals.

En l'apartat *a*) cal transformar els valors originals mitjançant l'IPC per a convertir tots els valors en els equivalents de l'any 2006. Aquesta operació s'anomena *deflactació* de la sèrie.

En l'apartat *b*) calcularem els índexs encadenats amb les dades transformades de l'apartat anterior.

L'apartat *c*) es calcula com en els exercicis 1 i 2. És recomanable utilitzar els valors primer i darrer de la llista de l'apartat *a*).

L'apartat *d*) inclou diferents càlculs que indiquem a continuació. Recordeu que per a fer previsions cap al futur fem la hipòtesi que les magnituds evolucionen al ritme de l'increment mitjà anual que agafem per a fer els càlculs.

Cal fer les previsions dels impostos en termes reals, multiplicant la darrera data coneguda en termes reals i multiplicant-la per l'increment mitjà anual tantes vegades com anys passaran.

Cal fer també les previsions de l'IPC dels anys venidors, aplicant a la darrera data publicada l'increment mitjà anual obtingut dels IPC dels anys considerats.

També necessitem convertir les previsions dels resultats anteriors en termes reals, a termes relatius o monedes corrents utilitzant els valors de l'IPC de l'any base i les estimacions de l'IPC dels anys venidors.

Exercici 8

En aquest exercici anem a calcular la pèrdua o guany de poder adquisitiu d'un salari del que coneixem els valors.

A més, s'han d'obtenir els valors de l'IPC dels mateixos anys de la web de l'INE.

Amb les dues llistes és necessari calcular els increments anuals mitjançant els índexs encadenats.

Recordar que per a determinar les variacions del poder adquisitiu cal fer aquesta operació:

$$\text{Guany o pèrdua de poder adquisitiu} = \frac{(1 + \Delta_{salari})}{(1 + \Delta_{IPC})}$$

Després es completa la taula amb la columna equivalent, calculada amb les dades inicial i final del període.

Acabarem amb aquesta taula completada, on ja hem indicat els increments anuals abans esmentats:

	2008	**2009**	**2010**	**2011**	**TOTAL**
Increment salarial	+1,4%	+7,6%	+1,6%	-3,2%	
Increment IPC	+4,1%	-0,3%	+1,8%	+2,1%	
Pèrdua o guany poder adquisitiu					

Exercici 9

Per a fer aquest exercici tan sols s'ha de fer el mateix esquema de l'exercici núm. 8.

Exercici 10

Per començar, cal fer un canvi de base i trobar tots els valors de l'IPC en la mateixa base. Emplenarem la taula amb els nous valors trobats, tal com es va fer a l'exercici 4.

	2000	2001	2002	2003	2004
IPC base 1992	131	135	139		
IPC base 2002			103	106	109

a) Calculeu l'increment anual, mitjà i total del preu de l'ordinador en termes reals.

Ja que l'anàlisi del preu de l'ordinador ens el demanen en termes reals, primer s'ha de fer la conversió del seu preu a moneda constant de l'any 2000 (deflactació de la sèrie de preus de l'ordinador). Podem ajudar-nos de les següents taules per a donar els resultats de cada any amb més claredat.

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)
2000	1300	1300
2001	1275	$1275 \cdot \dfrac{IPC_{1992}^{2000}}{IPC_{1992}^{2001}} =$ ....
2002	1250	$1250 \cdot \dfrac{IPC_{1992}^{2000}}{IPC_{1992}^{2002}} =$
2003	1100	
2004	950	

Per a calcular l'increment anual del preu de l'ordinador en termes reals, operem en la taula següent mitjançant índexs encadenats.

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)	Índex	Interpretació: Increment anual
2000	1300	1300	-----	
2001	1275	1237,2	$I_{2000}^{2001} = \frac{1237,2}{1300} = ...$	Ha disminuït un ....%
2002	1250	1178,06		
2003	1100	1007,34		
2004	950	846,02		

I finalment, amb les dades de la taula, calcularem l'increment total i mitjà del període.

b) Si seguim aquesta evolució, estimeu el preu que podria tenir l'ordinador l'any 2008.

Per a fer aquest apartat, considerem que ens demanen l'estimació en termes corrents del preu, suposant que no varia el comportament de l'IPC ni l'evolució del preu de l'ordinador en termes reals, que hem analitzat a l'apartat anterior.

Per a fer aquestes estimacions necessitem l'increment mitjà de les dues sèries i amb aquestes dades, podrem estimar el valor de l'IPC a l'any 2008.

Farem les mateixes operacions amb la sèrie dels preus dels ordinadors en termes reals i finalment, passarem el resultat a termes corrents en moneda del 2008.

Ajudes Tipus 2

Exercici 1

En l'apartat *a*), per acalcular els índex proposem emplenar aquesta taula:

	Nombre total alumnes matriculats	Índex Base 2005
Curs 2005/2006	12676	$I_{2005}^{2005} = 1$
Curs 2006/2007	12928	$I_{2005}^{2006} = \frac{12928}{12676} = 1{,}01988009$
Curs 2007/2008	13159	$I_{2005}^{2007} = \frac{13159}{12676} =$
Curs 2008/2009	13210	$I_{2005}^{2008} = \frac{13210}{12676} =$
Curs 2009/2010	13904	$I_{2005}^{2009} =$
Curs 2010/2011	14702	$I_{2005}^{2010} =$

Ara cal interpretar els resultats de la columna de la dreta.

En l'apartat *b*) ens demanen els índexs encadenats i també podem emplenar la taula següent:

	Nombre total alumnes matriculats	Índexs encadenats
Curs 2005/2006	12676	---
Curs 2006/2007	12928	$I_{2005}^{2006} = \frac{12928}{12676} = 1{,}01988009$
Curs 2007/2008	13159	$I_{2006}^{2007} = \frac{13159}{12928} =$
Curs 2008/2009	13210	$I_{2007}^{2008} =$
Curs 2009/2010	13904	$I_{2008}^{2009} =$
Curs 2010/2011	14702	$I_{2009}^{2010} =$

Interpretarem també els resultats de la columna de la dreta. Com que els índexs son tots majors que 1, indica que la sèrie sempre augmenta però a diferent ritme, segons l'any que analitzem.

En l'apartat *c*) ens demanen l'increment total del període $I_{2005}^{2010} = \dfrac{14702}{12676} = 1,16.$

I l'increment mitjà anual amb l'arrel quinta d'aquest quocient.

En l'apartat *d*) per a fer previsions, partim de la hipòtesi que l'increment mitjà anual que hem obtingut a l'apartat anterior serà una estimació de l'increment anual dels anys que estan per venir, i a partir de la darrera dada coneguda calcularem les quantitats d'alumnes que podrem esperar.

Així, per a estimar la quantitat d'alumnes que podem esperar que es matricule al curs 2011/2012, serà de 14702 · 1,03 = 15143 alumnes.

Amb el mateix raonament podem fer la resta d'estimacions.

Exercici 2

Aquest exercici és igual que l'exercici 1, però obtindrem alguns índexs per damunt d'1 i d'altres per sota. Deixem la interpretació per als lectors.

En l'apartat *a*) cal fer els càlculs que suggerim en la columna de la dreta de la taula següent:

	Nombre total visitants	Índex Base 2000
2000	10252799	$I_{2000}^{2000} = 1$
2001	10002517	$I_{2000}^{2001} = \dfrac{10002517}{10252799} =$
2002	9661493	$I_{2000}^{2002} = \dfrac{9661493}{10252799} = 0,94232736$
2003	10296382	$I_{2000}^{2003} = \dfrac{10296382}{10252799} = 1,00425084$
2004	11134880	$I_{2000}^{2004} = \dfrac{11134880}{10252799} =$
2005	10743480	$I_{2000}^{2005} = \dfrac{10743480}{10252799} = 1,04785825$

2006	10979470	$I_{2000}^{2006} = \dfrac{10979470}{10252799} =$
2007	10864738	$I_{2000}^{2007} =$
2008	10222818	$I_{2000}^{2008} =$
2009	9952606	$I_{2000}^{2009} =$

Cal fer la interpretació d'aquests resultats, especialment prestar atenció a la dels índexs menors que 1.

Per a determinar els índex encadenats de l'apartat *b*) també suggerim els càlculs a fer en la darrera columna de la taula següent:

	Nombre total visitants	Índexs encadenats
2000	10252799	
2001	10002517	$I_{2000}^{2001} = \dfrac{10002517}{10252799} = 0{,}97558891$
2002	9661493	$I_{2001}^{2002} = \dfrac{9661493}{10002517} =$
2003	10296382	$I_{2002}^{2003} = \dfrac{10296382}{9661493} =$
2004	11134880	$I_{2003}^{2004} = \dfrac{}{10296382} = 1{,}08143618$
2005	10743480	$I_{2004}^{2005} = \dfrac{10743480}{} = 0{,}96484919$
2006	10979470	$I_{2005}^{2006} = ---- = 1{,}02196588$
2007	10864738	$I_{2006}^{2007} =$
2008	10222818	$I_{2007}^{2008} =$
2009	9952606	$I_{2008}^{2009} =$

Deixem les interpretacions dels resultats al lector.

En l'apartat *c*) ens demanen l'increment total. Cal obtenir el següent índex I_{2000}^{2009} i amb el resultat fer l'arrel que ens permet esbrinar el valor de l'increment mitjà anual del –0,33%.

En l'apartat *d*) ens demanen previsions que calcularem a partir de la darrera data coneguda (9952606) i aplicant reiteradament el factor corresponent a l'increment mitjà anual del -0,33% (0,9967).

Exercici 3

Aquest exercici treballa els mateixos conceptes que els anteriors però començant per increments percentuals que permeten convertir-los en índexs encadenats directament. En la taula següent indiquem alguns valors i convidem a completar-la:

Any	Variacions del volum de vendes (%)	Índex encadenats
2006	-3,13	$I_{2005}^{2006} = 0{,}9687$
2007	-2,15	$I_{2006}^{2007} =$
2008	+2,12	$I_{2007}^{2008} = 1{,}0212$
2009	+3,15	$I_{2008}^{2009} =$
2010	+4,12	$I_{2009}^{2010} = 1{,}0412$
2011	+4,31	$I_{2010}^{2011} =$

Per a calcular els índexs de base 2005, multiplicarem els índexs anteriors (ací teniu determinats alguns d'ells. Calculeu-ne la resta).

$$I_{2005}^{2007} = I_{2005}^{2006} \cdot I_{2006}^{2007} = 0{,}9687 \cdot 0{,}9785 = 0{,}9479$$
$$I_{2005}^{2008} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} = 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 = 0{,}9680$$
$$I_{2005}^{2009} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} =$$
$$I_{2005}^{2010} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} =$$
$$I_{2005}^{2011} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} \cdot I_{2010}^{2011} =$$
$$= 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 \cdot 1{,}0412 \cdot 1{,}0431 = 1{,}0844$$

En l'apartat *b*) ens demanen l'increment total a partir dels factors dels índexs, que es el darrer càlcul que hem vist i així, l'increment mitjà anual és:

$$\sqrt[6]{0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 \cdot 1{,}0412 \cdot 1{,}0431} = \sqrt[6]{1{,}0844} = 1{,}0136$$

que podem interpretar com que l'augment total és equivalent a un increment anual constant de l'1,36%.

En l'aparat *c*) ens demanen previsions i cal reflexionar si aquest supòsit és possible.

Exercici 4

En aquest exercici ens presenten el procés de càlcul d'un canvi de base dels índexs de preus (IPC).

A tal fi cal considerar les dades que presentem en aquesta taula i plantejar proporcionalitats com les que mostrem a continuació per a emplenar les caselles ombrejades en gris.

	2002	2003	2004	2005	2006	2007	2008	2009	2010
IPC Base 2001	117,624	117,624	117,624	117,624	117,624				
IPC Base 2006	88,024	90,699			100	102,787	106,976	106,668	108,588

$$IPC_{2001}^{2002} = 103{,}538 \;\rightarrow\; IPC_{2006}^{2002} = \boldsymbol{x}$$

$$IPC_{2001}^{2006} = 117{,}624 \rightarrow IPC_{2006}^{2006} = 100$$

$$IPC_{2006}^{2002} = \frac{IPC_{2001}^{2002} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{103{,}538 \cdot 100}{117{,}624} = 103{,}538\frac{100}{117{,}624} =$$

$= 103{,}538 \cdot 0{,}8502 = 88{,}025$

$$IPC_{2006}^{2003} = \frac{IPC_{2001}^{2003} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{106{,}684 \cdot 100}{117{,}624} = 106{,}684\frac{100}{117{,}624} =$$

$= 106{,}684 \cdot 0{,}8502 = 90{,}699$

I així, fins a completar la taula.

És evident, que per a obtenir el valor de l'IPC de l'any «i» en base 2006 hem multiplicat l'IPC de l'any «i» en base 2001 per la fracció $\frac{IPC_{2006}^{2006}}{IPC_{2001}^{2006}}$ el valor de la qual hem ressenyat en roig i és el que denominarem «enllaç tècnic».

Encara que no es contempla a l'exercici, si dividim els valors de l'IPC de la segona taula per aquest «enllaç tècnic», també podríem obtenir els IPC del període 2006-2010 en base 2001.

Exercici 5

Com que a les ajudes de tipus 1 ja hem presentat les fórmules a emprar en cadascun dels casos i el significat de la notació, hi posarem un exemple fet de cadascuna:

Índex de preus per Laspeyres

$$L_{preus}{}_{2008}^{2009} = \frac{\sum_i p_{i2009} \cdot q_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{14 \cdot 100 + 8 \cdot 50 + 10 \cdot 20}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2000}{1800} = 1{,}111$$

Índex de quantitats per Laspeyres

$$L_{quantitats}{}_{2008}^{2009} = \frac{\sum_i q_{i2009} \cdot p_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{112 \cdot 12 + 65 \cdot 10 + 10 \cdot 5}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2044}{1800} = 1{,}136$$

Índex de preus per Paasche

$$P_{preus}{}_{2008}^{2009} = \frac{\sum_i p_{i2009} \cdot q_{i2009}}{\sum_i p_{i2008} \cdot q_{i2009}} = \frac{14 \cdot 112 + 8 \cdot 65 + 10 \cdot 10}{12 \cdot 112 + 10 \cdot 65 + 5 \cdot 10} = \frac{2188}{2044} = 1{,}070$$

Índex de quantitats per Paasche

$$P_{quantitats}{}_{2008}^{2009} = \frac{\sum_i p_{i2009} \cdot q_{i2009}}{\sum_i q_{i2008} \cdot p_{i2009}} = \frac{14 \cdot 112 + 8 \cdot 65 + 10 \cdot 10}{14 \cdot 100 + 8 \cdot 50 + 10 \cdot 20} = \frac{2188}{2000} = 1{,}094$$

Exercici 6

La primera operació permetrà fer el canvi a euros de la quantitat de compra: 96913,20 euros en moneda corrent de l'any 1998, i per a transformar-la en moneda de l'any 2006, farem l'operació següent:

$$96913{,}20 \cdot \frac{IPC_{2001}^{2006}}{IPC_{2001}^{1998}} = 96913{,}20 \cdot \frac{117{,}624}{91{,}223} = 124961{,}01€$$

Els valors de l'IPC es troben en la web de l'INE amb el conseqüent canvi de base com vam aprendre a l'exercici 4.

Com que a l'enunciat es diu que l'hem venut per 240000 euros, anem a calcular els beneficis en termes relatius, a partir del concepte de l'índex: $\frac{240000}{124961{,}01}$. Interpreteu-ne el resultat.

Exercici 7

Per a deflactar la sèrie (passar-la a termes reals de l'any 2006) hem de fer les operacions que es reflecteixen a la taula següent. A tal fi, consultarem els valors de l'IPC d'aquests anys. Prendrem les mitjanes anuals de l'IPC general.

Any	Import impost municipal (termes nominals)	Import impost municipal (termes reals 2006)
2006	503,24	503,24
2007	515,65	$515{,}65 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2007}} = 515{,}65 \cdot \frac{100}{102{,}787} = 501{,}67$
2008	536,73	$536{,}73 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2008}} = 536{,}73 \cdot \frac{100}{106{,}976} =$
2009	578,84	$578{,}84 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2009}} = 578{,}84 \cdot \frac{100}{106{,}668} =$
2010	584,42	$584{,}42 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2010}} = 584{,}42 \cdot \frac{100}{108{,}588} =$

En l'apartat *b*) ens demanen els índexs encadenats amb els valors en termes reals que presentem en la taula següent.

Any	Import impost municipal (termes nominals)	Import impost municipal (termes reals 2006)	Índex	
2006	503,24	503,24	-----	
2007	515,65	501,67	$I_{2006}^{2007}=\dfrac{501,67}{503,24}=0,997$	Ha disminuït un 0,3%
2008	536,73	501,73	$I_{2007}^{2008}=\dfrac{501,73}{501,67}$	Podem considerar que és constant
2009	578,84	542,66	$I_{2008}^{2009}=$	
2010	584,42	538,20	$I_{2009}^{2010}=$	Ha disminuït un .....%

En l'apartat *c*) ens demanen l'increment total i l'increment mitjà anual però disposem de les magnituds de la columna corresponent (tercera de la taula anterior).

Per a calcular l'increment total del període, interpretarem l'índex:

$$I_{2006}^{2010}=\frac{538,20}{503,24}$$

i per a determinar l'increment mitjà anual, calcularem l'arrel quarta del resultat anterior.

Per a abordar l'apartat *d*), i com que es tracta de fer estimacions, suposarem que els fenòmens evolucionaran amb el ritme que podem interpretar de l'increment mitjà anual de cadascuna.

Calcularem, en primer lloc, l'increment que correspon als IPC del anys 2006 al 2010 amb les dades de la web de l'INE. Hi hem obtingut un increment del 2,08% anual i amb aquesta dada i l'IPC de l'any 2010 estimarem els IPC dels anys 2011, 2012 i 2013.

També estimarem l'import de l'impost en termes reals d'aquests anys, aplicant-hi l'increment mitjà anual (1,7%) sobre el valor d'aquest import de l'any 2010, és a dir, 538,20 euros en termes reals de l'any 2006.

Com aquests resultats estan expressats en termes reals de l'any 2006, cal convertir-los a termes nominals, utilitzant els IPC que acabem d'estimar també.

Indiquem les operacions del primer resultat:

$$\text{Any 2011} \rightarrow 547,25\cdot\frac{IPC_{2006}^{2011}}{IPC_{2006}^{2006}}=547,25\cdot\frac{110,847}{100}=606,61$$

I la resta es calculen de la mateixa manera.

Exercici 8

En aquest exercici es tracta d'estudiar la pèrdua o guany de poder adquisitiu d'un salari.

Caldrà calcular els increments anuals dels salaris com indiquem en la taula següent:

Any	Nòmina mensual (€)	Índex	
2007	2034,75	------	-------
2008	2062,13	$I_{2007}^{2008}=\frac{2062,13}{2034,75}=1,014$	Ha augmentat 1,4%
2009	2218,61	$I_{2008}^{2009}=\frac{\quad}{2062,13}=1,076$	Ha augmentat ....%
2010	2253,67	$I_{2009}^{2010}=\frac{2253,67}{\quad}=$	Ha augmentat 1,6%
2011	2181,75	$I_{2010}^{2011}=----=0,968$	Ha disminuït ...%

També calcularem els increments anuals de l'IPC dels anys corresponents:

Any	IPC Base 2006	Índex	
2007	$IPC_{2006}^{2007}=102,787$	------	-------
2008	$IPC_{2006}^{2008}=106,976$	$I_{2007}^{2008}=\frac{106,976}{102,787}=1,041$	Ha augmentat 4,1%
2009	$IPC_{2006}^{2009}=106,668$	$I_{2008}^{2009}=\frac{106,668}{106,976}=$	Ha disminuït ---%
2010	$IPC_{2006}^{2010}=108,588$	$I_{2009}^{2010}=\frac{108,588}{\quad}=$	Ha augmentat 1,8 %
2011	$IPC_{2006}^{2011}=110,847$	$I_{2010}^{2011}=\frac{\quad}{108,588}=$	Ha augmentat ...%

Amb aquests resultats cal emplenar les caselles en gris de la taula següent:

	2008	2009	2010	2011
Increment salarial	1,014	1,076	1,016	0,968
Increment IPC	1,041	0,997	1,018	1,021
Pèrdua o guany poder adquisitiu	0,974			

Indiquem com obtenir el resultat que hem posat en la taula.

Any 2008 → $\frac{1{,}014}{1{,}041} = 0{,}974$, de la mateixa manera podem omplir la resta de caselles.

Amb els valors dels IPC i dels salaris podem obtenir l'increment total del període de cadascuna de les magnituds. Podem afegir els resultats a la columna de la dreta de la taula:

Indiquem tots els resultats obtinguts per comprovar:

	2008	2009	2010	2011	TOTAL
Increment salarial	+1,4%	+7,6%	+1,6%	-3,2%	+7,2%
Increment IPC	+4,1%	-0,3%	+1,8%	+2,1%	+7,8%
Pèrdua o guany poder adquisitiu	-2,6%	+7,9%	-0,2%	-5,2%	-0,6%

Exercici 9

Per a fer aquest exercici has de seguir les pautes de l'exercici núm. 8.

Exercici 10

Per començar cal fer un canvi de base i trobar tots els valors de l'IPC en la mateixa base. Emplenarem la taula, amb els nous valors trobats, com es va fer a l'exercici 4.

	2000	2001	2002	2003	2004
IPC base 1992	131	135	139		
IPC base 2002			103	106	109

Poden plantejar-se per proporcionalitat, com ací es pot veure per a l'any 2003:

$$IPC_{2002}^{2002} = 103 \rightarrow IPC_{2002}^{2003} = 106$$

$IPC_{1992}^{2002} = 139 \rightarrow IPC_{1992}^{2003} = \text{X}$ i calcular la dada desconeguda.

a) Calculeu l'increment anual, mitjà i total del preu de l'ordinador en termes reals.

Com que l'anàlisi del preu de l'ordinador ens el demanen en termes reals, cal fer primer la conversió del seu preu a moneda constant de l'any 2000 (deflactació de la sèrie de preus de l'ordinador) com podem veure en la taula següent;

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)
2000	1300	1300
2001	1275	$1275\cdot\frac{IPC_{1992}^{2000}}{IPC_{1992}^{2001}}=1275.\frac{131}{135}=1237{,}2$
2002	1250	$1250\cdot\frac{IPC_{1992}^{2000}}{IPC_{1992}^{2002}}=$
2003	1100	$1100\cdot\frac{IPC_{1992}^{2000}}{IPC_{1992}^{2003}}=$
2004	950	$950\cdot\frac{IPC_{1992}^{2000}}{IPC_{1992}^{2004}}=$

Per a calcular l'increment anual del preu de l'ordinador en termes reals, operem en la taula següent mitjançant índexs encadenats (intenta acabar-la).

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)	Índex	Interpretació: Increment anual
2000	1300	1300	-----	
2001	1275	1237,2	$I_{2000}^{2001}=\frac{1237{,}2}{1300}=0{,}9517$	Ha disminuït un 4,83%
2002	1250	1178,06		Ha disminuït un 4,78%
2003	1100	1007,34	$I_{2002}^{2003}=\frac{1007{,}34}{1178{,}06}=0{,}8551$	
2004	950	846,02		

Per a calcular l'increment total i mitjà del període, calcularem I_{2000}^{2004} i l'arrel corresponent d'índex 4.

b) Si seguim aquesta evolució, estimem el preu que podria tenir l'ordinador l'any 2008.

Per a fer aquest apartat, considerem que ens demanen l'estimació en termes corrents del preu, suposant que no varia el comportament de l'IPC ni l'evolució del preu de l'ordinador en termes reals que hem analitzat en l'apartat anterior.

Per a realitzar aquestes estimacions necessitem l'increment mitjà de les dues sèries i amb aquestes dades, podem estimar el valor de l'IPC a l'any 2998:

$$IPC_{1992}^{2008} = IPC_{1992}^{2004} \cdot 1{,}0294^4 = 147{,}10 \cdot 1{,}0294^4 = 165{,}18$$

Farem les mateixes operacions amb la sèrie del preus dels ordinadors en termes reals.

Preu ordinador per a l'any 2008 en termes reals del 2000 = $846{,}02 \cdot 0{,}8982^4$ = 550,65 euros.

I finalment, cal passar-los a termes corrents en moneda del 2008.

Solucions

Exercici 1

A continuació presentem el volum total d'alumnes matriculats a la Universitat Jaume I els darreres anys.

	Nombre total alumnes matriculats
Curs 2005/2006	12676
Curs 2006/2007	12928
Curs 2007/2008	13159
Curs 2008/2009	13210
Curs 2009/2010	13904
Curs 2010/2011	14702

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2005 (farà referència al curs 2005/2006). Interpreteu-ne els resultats.
b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà anual d'aquest període, a partir de les quantitats originals i a partir dels índexs encadenats.
d) Feu previsions per a la matrícula del curs 2011/2012 i 2012/2013 si considerem que no hi haurà canvis significatius en el seu comportament.

Solució

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2005 (farà referència al curs 2005/2006). Interpreteu-ne el resultat.

Agafarem com a data de referència, els 12676 alumnes matriculats el curs 2005/2006, i en la següent taula indicarem els quocients corresponents als índexs que volem calcular:

	Nombre total alumnes matriculats	Índex Base 2005
Curs 2005/2006	12676	$I_{2005}^{2005} = 1$
Curs 2006/2007	12928	$I_{2005}^{2006} = \frac{12928}{12676} = 1{,}01988009$
Curs 2007/2008	13159	$I_{2005}^{2007} = \frac{13159}{12676} = 1{,}0381035$
Curs 2008/2009	13210	$I_{2005}^{2008} = \frac{13210}{12676} = 1{,}04212685$
Curs 2009/2010	13904	$I_{2005}^{2009} = \frac{13904}{12676} = 1{,}09687599$
Curs 2010/2011	14702	$I_{2005}^{2010} = \frac{14702}{12676} = 1{,}1598296$

Per a interpretar aquestes dades considerarem que aquests quocients comparen dues magnituds en termes relatius i son quantitats que no vénen expressades en cap unitat sinó que podem interpretar-les com a percentatges. Així, si arrodonim els resultats amb dues xifres decimals podrem dir:

$I_{2005}^{2006} = 1{,}02$ ens indica que el nombre d'alumnes del curs 2006 és un 2% superior al de l'any 2005. També podríem expressar-los així $I_{2005}^{2006} = \frac{12928}{12676} \cdot 100 = 102\%$.

$I_{2005}^{2007} = 1{,}04$ ens indica que el nombre d'alumnes del curs 2007 és un 4% superior al de l'any 2005.

$I_{2005}^{2008} = 1{,}04$ ens indica que el nombre d'alumnes del curs 2008 és un 4% superior al de l'any 2005. Indica un cert estacionament en l'augment del nombre d'alumnes.

$I_{2005}^{2009} = 1{,}10$ ens indica que el nombre d'alumnes del curs 2009 és un 10% superior al de l'any 2005.

$I_{2005}^{2010} = 1{,}16$ ens indica que el nombre d'alumnes del curs 2010 és un 16% superior al de l'any 2005.

Aquests índexs donen idea de l'augment del nombre total d'alumnes de l'UJI, considerant sempre com a referència el nombre d'alumnes del curs 2005 que podrem considerar que serà del 100%.

b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.

Per a calcular aquests índexs, utilitzarem la taula següent, que podrem comparar-la amb la de l'apartat anterior:

	Nombre total alumnes matriculats	Índexs encadenats
Curs 2005/2006	12676	
Curs 2006/2007	12928	$I_{2005}^{2006} = \dfrac{12928}{12676} = 1{,}01988009$
Curs 2007/2008	13159	$I_{2006}^{2007} = \dfrac{13159}{12928} = 1{,}01786819$
Curs 2008/2009	13210	$I_{2007}^{2008} = \dfrac{13210}{13159} = 1{,}00387567$
Curs 2009/2010	13904	$I_{2008}^{2009} = \dfrac{13904}{13210} = 1{,}05253596$
Curs 2010/2011	14702	$I_{2009}^{2010} = \dfrac{14702}{13904} = 1{,}05739356$

Com es pot veure, en aquests índexs, la quantitat que prenem com a referència és el nombre d'alumnes matriculats l'any anterior, per la qual cosa podem denominar-los «índexs encadenats» i ens permetrà veure el creixement any rere any.

$I_{2005}^{2006} = 1{,}02$ ens indica que el nombre d'alumnes del curs 2006 és un 2% superior al de l'any 2005.

$I_{2006}^{2007} = 1{,}02$ ens indica que el nombre d'alumnes del curs 2007 és un 2% superior al de l'any 2006.

$I_{2007}^{2008} = 1{,}004$ ens indica que el nombre d'alumnes del curs 2008 és un 0,4% superior al de l'any 2007.

$I_{2008}^{2009} = 1{,}05$ ens indica que el nombre d'alumnes del curs 2009 és un 5% superior al de l'any 2008.

$I_{2009}^{2010} = 1{,}06$ ens indica que el nombre d'alumnes del curs 2010 és un 6% superior al de l'any 2009.

Aquests índexs donen idea de l'augment del nombre total d'alumnes de l'UJI, però detallant l'evolució per anys.

Calcula l'increment total i increment mitjà anual d'aquest període, a partir de les quantitats originals i a partir dels índexs encadenats.

Per a calcular l'increment total basant-nos en les quantitats originals, tan sols cal considerar el nombre d'alumnes dels primers i darrers cursos, per a comparar-los. Aquest concepte correspon al darrer índex calculat a l'apartat *a*). Així:

$I_{2005}^{2010} = \frac{14702}{12676} = 1{,}16$ que ja hem comentat que ens indica que els nombre d'alumnes s'ha incrementat un 16% en el període estudiat.

Per a obtenir l'increment mitjà anual, caldrà considerar l'arrel amb un índex que ve donat pel nombre d'increments que conté el període estudiat. Notem que correspon al nombre d'anys menys un. Així al nostre exercici:

$$\sqrt[5]{\frac{14702}{12676}} = \sqrt[5]{1{,}1598296} = 1{,}03$$

i podrem interpretar que aquest creixement total del 16%, seria equivalent a un creixement constant anual del 3%.

Ressenyem que en aquests càlculs ja fets hem utilitzat el nombre d'alumnes del primer i darrer curs que considerem en el període.

En altres magnituds, podem no conèixer aquestes quantitats però si els increments anuals o índexs encadenats que nosaltres hem obtingut a l'apartat *b*). En aquest cas també podrem calcular l'increment total i l'increment mitjà anual. Vegem:

$$I_{2005}^{2010} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} =$$

$$= \frac{12928}{12676} \cdot \frac{13159}{12928} \cdot \frac{13210}{13159} \cdot \frac{13904}{13210} \cdot \frac{14702}{13904} = \frac{14702}{12676} =$$

$= 1{,}01988009 \cdot 1{,}01786819 \cdot 1{,}00387567 \cdot 1{,}05253596 \cdot 1{,}05739356 = 1{,}1598296$ que ens permet interpretar que en el període considerat el nombre d'alumnes ha augmentat un16%.

Per a determinar l'increment mitjà, calcularem doncs l'arrel quinta del producte dels índexs (l'índex de l'arrel coincideix amb el nombre d'índexs que formen el producte).

$$\sqrt[5]{I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010}} =$$

$$\sqrt[5]{1{,}01988009 \cdot 1{,}01786819 \cdot 1{,}00387567 \cdot 1{,}05253596 \cdot 1{,}05739356} = \sqrt[5]{1{,}1598296} = 1{,}03$$

que interpretarem com un increment anual constant del 3%.

Feu previsions per a la matrícula del curs 2011/2012 i 2012/2013 si considerem que no hi haurà canvis significatius en el seu comportament.

Per a fer previsions, cal partir de la hipòtesi que l'increment mitjà anual que hem obtingut en l'apartat anterior podria ser una estimació de l'increment anual dels anys que estan per venir i, que a partir de la darrera dada coneguda, calcularem les quantitats d'alumnes que hi esperem.

Així, per a estimar la quantitat d'alumnes que podem esperar que es matricule al curs 2011/2012, serà de alumnes.

I per a estimar la quantitat d'alumnes que podrem esperar per al curs 2012/2013 cal considerar que passaran dos anys des de la data de la darrera quantitat d'alumnes real $14702 \cdot 1{,}03^2 = 15507$ alumnes.

Notem que per previsions o estimacions partirem de la darrera dada real i la multiplicarem per l'increment mitjà anual que hem calculat prèviament, i elevant-lo al nombre d'anys o períodes que estan per venir.

Exercici 2

En la següent taula mostra les dades de l'INI que fan referència al total de visitants als parcs nacionals d'Espanya, els anys que fem referència.

Natura i biodiversidad	
Zones protegides	
Nombre de visitants per parcs nacionals i període	
Unitats: nombre de persones	
	Total
2000	10252799
2001	10002517
2002	9661493
2003	10296382
2004	11134880
2005	10743480
2006	10979470
2007	10864738
2008	10222818
2009	9952606
Font:Ministeri de Medi Ambient i Medi Rural i Marí. Xarxa de Parcs Naturals Copyright INE 2011	

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2000 i interpreteu-ne els resultats.
b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà anual d'aquest període, a partir de les quantitats originals i a partir dels índexs encadenats.
d) Feu previsions del nombre de visitants dels parcs considerats per als anys 2010, 2011 i 2012, si considerem que no hi haurà canvis significatius en el comportament de l'afluència.

Font: INE

Solució

a) Calculeu els índexs per a cada any, prenent com a any de referència el 2000 i interpreteu-ne els resultats.

Agafarem com a data de referència, els 10252799 visitants que van recòrrer els parcs naturals d'Espanya en la seua totalitat l'any 2000, i en la taula següent indicarem els quocients corresponents als índexs que volem calcular:

	Nombre total visitants	Índex Base 2000
2000	10252799	$I_{2000}^{2000} = 1$
2001	10002517	$I_{2000}^{2001} = \frac{10002517}{10252799} = 0{,}97558891$
2002	9661493	$I_{2000}^{2002} = \frac{9661493}{10252799} = 0{,}94232736$
2003	10296382	$I_{2000}^{2003} = \frac{10296382}{10252799} = 1{,}00425084$
2004	11134880	$I_{2000}^{2004} = \frac{11134880}{10252799} = 1{,}08603319$
2005	10743480	$I_{2000}^{2005} = \frac{10743480}{10252799} = 1{,}04785825$
2006	10979470	$I_{2000}^{2006} = \frac{10979470}{10252799} = 1{,}07087538$
2007	10864738	$I_{2000}^{2007} = \frac{10864738}{10252799} = 1{,}05968507$
2008	10222818	$I_{2000}^{2008} = \frac{10222818}{10252799} = 0{,}99707582$
2009	9952606	$I_{2000}^{2009} = \frac{9952606}{10252799} = 0{,}97072087$

Per a interpretar aquestes dades considerarem que aquests quocients comparen dues magnituds en termes relatius i podem interpretar-les com a percentatges. A diferència de l'apartat anterior, aquesta magnitud disminueix i creix segons de quin període siguen les dades. Així, si arrodonim els resultats amb dues xifres decimals podrem dir:

$I_{2000}^{2001} = 0{,}98$ ens indica que el nombre de visitants ha disminuït un 2% de l'any 2000 al 2001.

$I_{2000}^{2002} = 0{,}94$ ens indica que el nombre de visitants ha disminuït un 6% de l'any 2000 al 2002.

$I_{2000}^{2003} = 1{,}004$ ens indica que el nombre de visitants ha augmentat un 0,4% de l'any 2000 al 2003.

$I_{2000}^{2004} = 1{,}09$ ens indica que el nombre de visitants ha augmentat un 9% de l'any 2000 al 2004.

$I_{2000}^{2005} = 1{,}05$ ens indica que el nombre de visitants ha augmentat un 5% de l'any 2000 al 2005.

$I_{2000}^{2006} = 1{,}07$ ens indica que el nombre de visitants ha augmentat un 7% de l'any 2000 al 2006.

$I_{2000}^{2007} = 1{,}06$ ens indica que el nombre de visitants ha augmentat un 6% de l'any 2000 al 2007.

$I_{2000}^{2008} = 0{,}997$ ens indica que el nombre de visitants ha disminuït un 0,3% de l'any 2000 al 2008.

$I_{2000}^{2009} = 0{,}97$ ens indica que el nombre de visitants ha disminuït un 3% de l'any 2000 al 2009.

Aquests índexs donen idea de les variacions en el nombre de visitants als parcs nacionals d'Espanya, malgrat que tan sols consideren els valors del començament i l'acabament del període referit i no s'hi reflecteixen les fluctuacions dintre del període.

b) Calculeu els índexs encadenats d'aquesta sèrie. Interpreteu-ne els resultats.

Per a calcular aquests índexs, utilitzarem la taula següent, que podrem comparar-la amb la de l'apartat anterior:

	Nombre total visitants	Índexs encadenats
2000	10252799	
2001	10002517	$I_{2000}^{2001} = \dfrac{10002517}{10252799} = 0{,}97558891$
2002	9661493	$I_{2001}^{2002} = \dfrac{9661493}{10002517} = 0{,}96590618$
2003	10296382	$I_{2002}^{2003} = \dfrac{10296382}{9661493} = 1{,}06571334$
2004	11134880	$I_{2003}^{2004} = \dfrac{11134880}{10296382} = 1{,}08143618$
2005	10743480	$I_{2004}^{2005} = \dfrac{10743480}{11134880} = 0{,}96484919$
2006	10979470	$I_{2005}^{2006} = \dfrac{10979470}{10743480} = 1{,}02196588$
2007	10864738	$I_{2006}^{2007} = \dfrac{10864738}{10979470} = 0{,}98955032$
2008	10222818	$I_{2007}^{2008} = \dfrac{10222818}{10864738} = 0{,}94091712$
2009	9952606	$I_{2008}^{2009} = \dfrac{9952606}{10222818} = 0{,}97356776$

Com es pot veure en aquests índexs, la quantitat que prenem com a referència és el nombre de visitants dels parcs de l'any anterior, per la qual cosa podem denominar-los «índexs encadenats» i ens permetrà veure el creixement any rera any.

$I_{2000}^{2001} = 0{,}98$ ens indica que el nombre de visitants ha disminuït un 2% de l'any 2000 al 2001.

$I_{2001}^{2002} = 0{,}97$ ens indica que el nombre de visitants ha disminuït un 3% de l'any 2001 al 2002.

$I_{2002}^{2003} = 1{,}07$ ens indica que el nombre de visitants ha augmentat un 7% de l'any 2002 al 2003.

$I_{2003}^{2004} = 1{,}08$ ens indica que el nombre de visitants ha augmentat un 8% de l'any 2003 al 2004.

$I_{2004}^{2005} = 0{,}96$ ens indica que el nombre de visitants ha disminuït un 4% de l'any 2004 al 2005.

$I_{2005}^{2006} = 1{,}02$ ens indica que el nombre de visitants ha augmentat un 2% de l'any 2005 al 2006.

$I_{2006}^{2007} = 0{,}99$ ens indica que el nombre de visitants ha disminuït un 1% de l'any 2006 al 2007.

$I_{2007}^{2008} = 0{,}94$ ens indica que el nombre de visitants ha disminuït un 6% de l'any 2007 al 2008.

$I_{2008}^{2009} = 0{,}97$ ens indica que el nombre de visitants ha disminuït un 3% de l'any 2008 al 2009.

Aquests índexs donen idea de l'augment del nombre total de visitants en tot el període detallant l'evolució per anys, per la qual cosa podem diferenciar els anys en què el nombre ha augmentat i en els que ha disminuït.

c) Calculeu l'increment total i l'increment mitjà anual d'aquest període, a partir de les quantitats originals i a partir dels índexs encadenats.

Per a calcular l'increment total basant-nos en les quantitats originals, tan sols cal considerar el nombre de visitants del primer i darrer any del període a analitzar. Aquest concepte correspon al darrer índex calculat en l'apartat *a*). Així:

0,97 ens indica que si comparem les visites de l'any 2009 amb les de l'any 2000, veurem que han disminuït en un 3%.

Per a obtenir l'increment mitjà anual, caldrà considerar l'arrel amb un índex 9 que ve donat pel nombre d'increments que conté el període estudiat. Notem que correspon al nombre d'anys o dades menys un. Així, amb les dades de l'exercici:

$$\sqrt[9]{\frac{9952606}{10252799}} = \sqrt[9]{0{,}97072087} = 0{,}9967$$

que podrem interpretar com que l'evolució total de disminució del 3% és equivalent a una disminució anual del 0,33%. Direm, per tant, que l'increment anual mitjà es del -0,33%.

Ressenyem que en aquests càlculs que hem fet, només hem utilitzat el nombre de visitants dels parcs dels anys 2000 i 2009, tan sols.

Vegem també com es poden calcular aquests mateixos increments, totals i mitjà, basant-nos en els índexs encadenats que hem obtingut en l'apartat *b*).

$$I_{2000}^{2009} = I_{2000}^{2001} \cdot I_{2001}^{2002} \cdot I_{2002}^{2003} \cdot I_{2003}^{2004} \cdot I_{2004}^{2005} \cdot I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} =$$

$$= 0{,}97558891 \cdot 0{,}96590618 \cdot 1{,}06571334 \cdot 1{,}08143618 \cdot 0{,}96484919 \cdot 1{,}02196588 \cdot 0{,}98955032 \cdot 0{,}94091712 \cdot 0{,}97356776 = 0{,}97072087$$

Per a calcular l'increment mitjà, avaluarem l'arrel novena del producte dels índexs (l'índex de l'arrel coincideix amb el nombre d'índexs que formen el producte).

$$\sqrt[9]{I_{2000}^{2001} \cdot I_{2001}^{2002} \cdot I_{2002}^{2003} \cdot I_{2003}^{2004} \cdot I_{2004}^{2005} \cdot I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009}} =$$

$$\sqrt[9]{0{,}9756 \cdot 0{,}9660 \cdot 1{,}0657 \cdot 1{,}0814 \cdot 0{,}9648 \cdot 1{,}0220 \cdot 0{,}9896 \cdot 0{,}9409 \cdot 0{,}9736} =$$

$$\sqrt[9]{0{,}9707} = 0{,}9967$$

que interpretarem com un disminució anual constant del 0,33%.

d) Feu previsions del nombre de visitants dels parcs considerats per als anys 2010, 2011 i 2012, si considerem que no hi haurà canvis significatius en el comportament de l'afluència.

Per a fer previsions, cal partir de la hipòtesi que l'increment mitjà anual que hem obtingut a l'apartat anterior podria ser una estimació de l'increment anual dels anys que estan per venir i, a partir de la darrera dada coneguda, calcularem les quantitats de visitants que podrem esperar per als anys 2010, 2011 i 2012.

$9952606 \cdot 0{,}9967 = 9919762$ visitants
$9952606 \cdot 0{,}9967^2 = 9887027$ visitants
$9952606 \cdot 0{,}9967^3 = 9854400$ visitants

Notem que per previsions o estimacions partirem de la darrera dada real i la multiplicarem per l'increment mitjà anual que hem calculat prèviament, i elevant-lo al nombre d'anys o períodes que estan per venir.

Exercici 3

A continuació presentem les variacions percentuals del volum de vendes de certa superfície comercial, els darrers anys.

Any	**Variacions del volum de vendes (%)**
2006	–3,13
2007	–2,15
2008	+2,12
2009	+3,15
2010	+4,12
2011	+4,31

a) Calculeu els índexs de les vendes de cada any, prenent com a referència l'any 2005 i els índexs encadenats.
b) Calculeu la variació o increment mitjà anual i total de les vendes en aquest període.
c) Estimeu les vendes dels dos anys següents si suposem que no hi hauran canvis significatius en el comportament de les vendes en aquests anys.

Solució

a) Calculeu els índexs de les vendes de cada any, prenent com a referència l'any 2005 i els índexs encadenats.

Cal veure que les dades d'aquest exercici es diferencien dels dos anteriors, ja que en aquest cas són increments percentuals anuals, per la qual cosa, les dades de la taula es podran «traduir» i convertir en índexs encadenats, tal com indiquem en la taula següent:

Any	Variacions del volum de vendes (%)	Índexs encadenats
2006	–3,13	$I_{2005}^{2006} = 0{,}9687$
2007	–2,15	$I_{2006}^{2007} = 0{,}9785$
2008	+2,12	$I_{2007}^{2008} = 1{,}0212$
2009	+3,15	$I_{2008}^{2009} = 1{,}0315$
2010	+4,12	$I_{2009}^{2010} = 1{,}0412$
2011	+4,31	$I_{2010}^{2011} = 1{,}0431$

Podem veure com calculem aquests índexs. A tal fi, li sumem o restem a 1 l'increment, i així convertim el tant per cent en tant per u:

$$I_{2005}^{2006} = 1 - 0.0313 = 0{,}9687$$

$$I_{2006}^{2007} = 1 - 0{,}0215 = 0{,}9785$$

$$I_{2007}^{2008} = 1 + 0{,}0212 = 1{,}0212$$

$$I_{2008}^{2009} = 1 + 0{,}0315 = 1{,}0315$$

I consegüentment amb la resta de valors de la columna de la dreta de la taula.

Per a calcular els índexs de base 2005, multiplicarem els índexs anteriors:

$$I_{2005}^{2007} = I_{2005}^{2006} \cdot I_{2006}^{2007} = 0{,}9687 \cdot 0{,}9785 = 0{,}9479$$

$$I_{2005}^{2008} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} = 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 = 0{,}9680$$

$$I_{2005}^{2009} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} = 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 = 0{,}9985$$

$$I_{2005}^{2010} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} =$$

$$= 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 \cdot 1{,}0412 = 1{,}0396$$

$$I_{2005}^{2011} = I_{2005}^{2006} \cdot I_{2006}^{2007} \cdot I_{2007}^{2008} \cdot I_{2008}^{2009} \cdot I_{2009}^{2010} \cdot I_{2010}^{2011} =$$

$$= 0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 \cdot 1{,}0412 \cdot 1{,}0431 = 1{,}0844$$

Presentarem tots els índex en la següent taula:

Any	Variacions del volum de vendes (%)	Índexs encadenats	Índex Base 2005
2006	–3,13	$I_{2005}^{2006} = 0{,}9687$	$I_{2005}^{2006} = 0{,}9687$
2007	–2,15	$I_{2006}^{2007} = 0{,}9785$	$I_{2005}^{2007} = 0{,}9479$
2008	+2,12	$I_{2007}^{2008} = 1{,}0212$	$I_{2005}^{2008} = 0{,}9680$
2009	+3,15	$I_{2008}^{2009} = 1{,}0315$	$I_{2005}^{2009} = 0{,}9985$
2010	+4,12	$I_{2009}^{2010} = 1{,}0412$	$I_{2005}^{2010} = 1{,}0396$
2011	+4,31	$I_{2010}^{2011} = 1{,}0431$	$I_{2005}^{2011} = 1{,}0844$

Si interpretem els índexs de la darrera columna podrem veure l'increment global en el volum de vendes de la superfície, en la casella inferior. Aquest índex indica un augment del +8,44% en el període analitzat.

b) Calculeu l'increment o variació mitjana anual i total de les vendes en aquest període.

Per a calcular l'increment total de les vendes tan sols cal interpretar l'índex $I_{2005}^{2011} = 1{,}0844$ que ens indica que les vendes han augmentat un 8,44% en el període analitzat.

Podem recordar que aquest índex l'hem calculat multiplicant els índexs encadenats que hem «construït» amb les dades dels percentatges de l'enunciat.

Per a determinar l'increment mitjà anual, calculem l'arrel d'índex 6, ja que fem una mitjana geomètrica amb els índexs de cada any. També podem partir de l'índex que representa l'increment total del període.

Així, l'increment mitjà anual és

$$\sqrt[6]{0{,}9687 \cdot 0{,}9785 \cdot 1{,}0212 \cdot 1{,}0315 \cdot 1{,}0412 \cdot 1{,}0431} = \sqrt[6]{1{,}0844} = 1{,}0136$$

que podem interpretar com que l'augment total és equivalent a un increment anual constant de l'1,36%.

d) Estimeu les vendes dels dos anys següents si suposem que no hi haurà canvis significatius en el comportament de les vendes en aquests anys.

No podem estimar les vendes perquè no coneixem la magnitud de les vendes d'algun any per agafar-lo de referència i a partir d'ell calcular les vendes de l'any que ens interessa.

Exercici 4

A continuació presentem els valors de l'índex de preus al consum (IPC), que podem consultar a la pàgina de l'INE, que fa referència a les dades en base al 2001 i 2006. Per raons que caldrà estudiar en la teoria, en certs moments s'ha de fer un canvi en l'any de referència i es comença a obtenir la nova sèrie de l'IPC, començat de nou amb el valor 100. Direm que hi ha hagut un «canvi de base».

Sovint, com podràs veure en exercicis posteriors, cal utilitzar en un mateix càlcul el valor de l'IPC d'anys que corresponen a períodes de bases diferents, i necessitarem treballar amb tots els valors de l'IPC referits a una mateixa base. Aquestes dades les podràs trobar fàcilment en la pàgina web de l'INE, però en aquest exercici veurem com es calculen els valors de les caselles que estan ombrejades en gris.

En primer lloc presentem la taula dels valors de l'IPC des de l'any 2002 al 2006 en base 2001.

Índex de preus de consum	
Mitjanes anuals. Base 2001	
Nacional per general i Grups COICOP	
Unitats: Índex i taxes	
	General
	Mitjana anual
2006	117,624
2005	113,63
2004	109,927
2003	106,684
2002	103,538

I a continuació, les dades dels valors de l'IPC des de l'any 2006 al 2010 en base 2006, encara que estan afegits els valors de les caselles grisses que corresponen als valors obtinguts a «posteriori» per facilitar els treballs de càlcul referits a períodes de diferents bases.

Índex de preus de consum	
Mitjanes anuals. Base 2006	
Índexs nacionals: general i de grups COICOP	
Unitats: Base 2006=100	
	General
	Media anual
2010	108,588
2009	106,668
2008	106,976
2007	102,787
2006	100
2005	96,604
2004	93,456
2003	90,699
2002	88,024

Expliqueu com s'han obtingut les dades de les caselles ombrejades en gris, esbrinant el valor de l'enllaç.

Font: INE

Solució

Aquest procés que denominem «canvi de base» tan sols és la transformació dels valors de l'IPC per assegurar-nos la proporcionalitat en la cadena dels valors dels períodes anteriors al moment en que per raons que, no vénen al cas, l'INE realitza aquesta actualització, i per tant, comença una nova sèrie de valors partint del 100.

Veiem que el càlcul és tan sols la resolució d'una proporcionalitat (regla de tres) on hi ha termes fixes que ens permetran trobar el valor de l'enllaç.

A la primera taula tenim els valors de l'IPC corresponents al període 2002-2006 en base 2001. Però arribat l'any 2006 es decideix un canvi de base i com es pot veure a la segona taula, tenim una nova sèrie de dades que comença en l'any 2006 amb un 100. En alguna situació que veurem en exercicis posteriors necessitem utilitzar l'IPC dels dos períodes, per la qual cosa, cal conèixer totes les dades referides a la mateixa base per no trencar la continuïtat de la seqüència que reflecteix el comportament dels preus al consum i, conseqüentment es converteix en un dels principals indicadors de la inflació i dels esdeveniments econòmics d'un país.

Vegem, doncs, com es calculen els valors de les caselles ombrejades en gris. Són els valors dels antics IPC en la nova base. Per al seu càlcul, tan sols cal plantejar les dades implicades per a assegurar-nos la proporcionalitat real en l'evolució dels preus.

Heus-ne ací les dades de l'any 2002:

Per calcular la dada desconeguda

$$IPC_{2006}^{2002} = \frac{IPC_{2001}^{2002} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{103{,}538 \cdot 100}{117{,}624} = 103{,}538 \frac{100}{117{,}624} = 103{,}538 \cdot 0{,}8502 =$$
$$= 88{,}025$$

De la mateixa manera, la resta d'anys repetim el procés:

$$IPC_{2006}^{2003} = \frac{IPC_{2001}^{2003} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{106{,}684 \cdot 100}{117{,}624} = 106{,}684 \frac{100}{117{,}624} = 106{,}684 \cdot 0{,}8502 =$$
$$= 90{,}699$$

$$IPC_{2006}^{2004} = \frac{IPC_{2001}^{2004} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{109{,}927 \cdot 100}{117{,}624} = 109{,}927 \frac{100}{117{,}624} = 109{,}927 \cdot 0{,}8502 =$$
$$= 93{,}456$$

$$IPC_{2006}^{2005} = \frac{IPC_{2001}^{2005} \cdot IPC_{2006}^{2006}}{IPC_{2001}^{2006}} = \frac{113{,}63 \cdot 100}{117{,}624} = 113{,}63 \frac{100}{117{,}624} = 113{,}63 \cdot 0{,}8502 = 96{,}604$$

És evident, que per a obtenir el valors de l'IPC de l'any «i» en base 2006 hem multiplicat l'IPC de l'any «i» en base 2001 per la fracció $\dfrac{IPC_{2006}^{2006}}{IPC_{2001}^{2006}}$ el valor de la qual hem ressenyat en roig i és el que denominarem «enllaç tècnic».

Encara que no es contempla a l'exercici, si dividim els valors de l'IPC de la segona taula per aquest «enllaç tècnic», també podríem obtenir els IPC del període 2006-2010 en base 2001.

La taula completa quedaria:

	2002	2003	2004	2005	2006	2007	2008	2009	2010
IPC Base2001	103,538	106,684	109,927	113,63	117,624	120,902	125,829	125,467	127,726
IPC Base2006	88,024	90,699	93,456	96,604	100	102,787	106,976	106,668	108,588

Exercici 5

Calcular els índexs de preus i quantitats dels articles A, B i C mitjançant la fórmula de Laspeyres i Paasche, dels anys 2008, 2009 i 2010 en funció de l'any 2008, utilitzant les dades de les següents taules on estan indicades les quantitats q_i i preus p_i que cal conèixer.

	2008		**2009**		**20010**	
	Preu p_i	Quantitat q_i	Preu p_i	Quantitat q_i	Preu p_i	Quantitat q_i
Art. A	12	100	14	112	15	115
Art. B	10	50	8	65	7	72
Art. C	5	20	10	10	15	5

Solució

Índex de preus per Laspeyres

Per a calcular aquests índexs, començarem per conèixer i deduir la fórmula a emprar. Hem reduït el seu càlcul a tres articles però no oblidem que aquest càlcul s'estén a la totalitat d'articles representatius del consum de les famílies en un país (veure ECPF).

$$L_{preus\,0}^{\;t} = \frac{\sum_i \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_i p_{i0} \cdot q_{i0}} = \frac{\sum_i p_{it} \cdot q_{i0}}{\sum_i p_{i0} \cdot q_{i0}}$$

on *t* és l'any actual i *0* serà l'any que prendrem com a referència en la comparació. Si es tracta d'índexs encadenats podríem nomenar-los anys *t-1* i *t*.

p_{it} serà el preu de l'article i *a* l'any *t*
p_{i0} serà el preu de l'article i *a* l'any 0
q_{it} serà la quantitat de l'article i *a* l'any *t*
q_{i0} serà la quantitat de l'article i *a* l'any 0

és una mitjana ponderada on el «pes» de cada article $p_{i0} \cdot q_{i0}$ és el valor de l'article en la «cistella de la compra» de l'any de referència i romandrà constant al llarg del període mentre no es canvie la base. Un inconvenient d'aquest mètode és que si la importància dels articles en els hàbits de consum canvia molt, aquests coeficients queden desfasats.

Així:

$$L_{preus\,2008}^{\;2008} = 1$$

$$L_{preus\,2008}^{\;2009} = \frac{\sum_i p_{i2009} \cdot q_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{14 \cdot 100 + 8 \cdot 50 + 10 \cdot 20}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2000}{1800} = 1{,}111$$

$$L_{preus\,2008}^{\;2010} = \frac{\sum_i p_{i2010} \cdot q_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{15 \cdot 100 + 7 \cdot 50 + 15 \cdot 20}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2150}{1800} = 1{,}194$$

Es pot comprovar que el denominador no varia i tan sols cal actualitzar els preus dels articles en el període nou a comparar. Aquest és un gran avantatge d'aquesta fórmula.

Índex de quantitats per Laspeyres

En aquest cas estudiarem l'evolució de les quantitats demandades i per a la ponderació s'utilitzaran els mateixos coeficients de l'apartat anterior $p_{i0} \cdot q_{i0}$.

$$L_{quantitats\,0}^{\;t} = \frac{\sum_i \frac{q_{it}}{q_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_i p_{i0} \cdot q_{i0}} = \frac{\sum_i q_{it} \cdot p_{i0}}{\sum_i p_{i0} \cdot q_{i0}}$$

Així:

$$L_{quantitats}{}_{2008}^{2008} = 1$$

$$L_{quantitats}{}_{2008}^{2009} = \frac{\sum_i q_{i2009} \cdot p_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{112 \cdot 12 + 65 \cdot 10 + 10 \cdot 5}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2044}{1800} = 1{,}136$$

$$L_{quantitats}{}_{2008}^{2010} = \frac{\sum_i p_{i2010} \cdot q_{i2008}}{\sum_i p_{i2008} \cdot q_{i2008}} = \frac{115 \cdot 12 + 72 \cdot 10 + 50 \cdot 5}{12 \cdot 100 + 10 \cdot 50 + 5 \cdot 20} = \frac{2350}{1800} = 1{,}306$$

Índex de preus per Paasche

$$P_{preus}{}_0^t = \frac{\sum_i \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{it}}{\sum_i p_{i0} \cdot q_{it}} = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i p_{i0} \cdot q_{it}}$$

És una mitjana ponderada on el «pes» de cada article $p_{i0} \cdot q_{it}$ intenta millorar la proposta de Laspeyres, evitant en certa manera el desfasament, ja que recull la importància de l'article en considerar la quantitat en el període a comparar.

Així:

$$P_{preus}{}_{2008}^{2008} = 1$$

$$P_{preus}{}_{2008}^{2009} = \frac{\sum_i p_{i2009} \cdot q_{i2009}}{\sum_i p_{i2008} \cdot q_{i2009}} = \frac{14 \cdot 112 + 8 \cdot 65 + 10 \cdot 10}{12 \cdot 112 + 10 \cdot 65 + 5 \cdot 10} = \frac{2188}{2044} = 1{,}070$$

$$P_{preus}{}_{2008}^{2010} = \frac{\sum_i p_{i2010} \cdot q_{i2010}}{\sum_i p_{i2008} \cdot q_{i2010}} = \frac{15 \cdot 115 + 7 \cdot 72 + 15 \cdot 50}{12 \cdot 115 + 10 \cdot 72 + 5 \cdot 50} = \frac{2979}{2350} = 1{,}268$$

Índex de quantitats per Paasche

$$P_{quantitats}{}_0^t = \frac{\sum_i \frac{q_{it}}{q_{i0}} \cdot q_{i0} \cdot p_{it}}{\sum_i q_{i0} \cdot p_{it}} = \frac{\sum_i p_{it} \cdot q_{it}}{\sum_i q_{i0} \cdot p_{it}}$$

Aquesta proposta, com que analitza l'evolució de les quantitats, considera com a coeficient $q_{i0} \cdot p_{it}$ que indica el «pes» de cada article, el preu de l'any *t* per a actualitzar la importància de l'article.

Així:

$$P_{quantitats}{}_{2008}^{2008} = 1$$

$$P_{quantitats}{}_{2008}^{2009} = \frac{\sum_i p_{i2009} \cdot q_{i2009}}{\sum_i q_{i2008} \cdot p_{i2009}} = \frac{14 \cdot 112 + 8 \cdot 65 + 10 \cdot 10}{14 \cdot 100 + 8 \cdot 50 + 10 \cdot 20} = \frac{2188}{2000} = 1{,}094$$

$$P_{quantitats}{}_{2008}^{2010} = \frac{\sum_i p_{i2010} \cdot q_{i2010}}{\sum_i q_{i2008} \cdot p_{i2010}} = \frac{15 \cdot 115 + 7 \cdot 72 + 15 \cdot 50}{15 \cdot 100 + 7 \cdot 50 + 15 \cdot 20} = \frac{2979}{2150} = 1{,}386$$

Com es pot veure, en aquests índexs, en cada un calculat per les fórmules de Paasche, cal determinar de nou sempre tant el numerador com el denominador de cada fracció. Aquesta diferència que ens pot semblar irrellevant per a tres articles, no ho sembla igual per a la gran quantitat de dades que cal treballar per al càlcul de l'IPC i amb els recursos tecnològics de temps enrere.

Exercici 6

Suposem que vam comprar un habitatge per 16125000 de pessetes al desembre de 1998 i l'hem venut al desembre del 2006 per un valor de 240000 euros. Esbrineu el percentatge de beneficis o pèrdues que hem tingut en l'operació.

Nota: Per a fer les operacions consultarem els valors de l'IPC que necessitem a la pàgina web de l'INE. www.ine.es (seria interessant calcular aquest increment amb l'IPC general i amb l'IPC del grup de l'habitatge).

El canvi de moneda que considerarem és 1 € = 166,386 PTA

Solució

Per començar a comparar caldrà treballar amb una única moneda. Decidim treballar en euros. És obvi que el resultat en termes relatius o percentatges, no varia si treballem en pessetes.

Per a transformar 16125000 de pessetes a euros utilitzarem el canvi que proposa la nota (*1 € = 166,386 PTA)* per la qual cosa, 16125000 / 166,386 = 96913,20 euros en termes corrents de desembre de 1998. Per a esbrinar quin seria el seu valor equivalent en termes corrents de l'any 2006, caldria fer la transformació següent:

$$96913{,}20 \cdot \frac{IPC_{2001}^{2006}}{IPC_{2001}^{1998}} = 96913{,}20 \cdot \frac{117{,}624}{91{,}223} = 124961{,}01 \text{ €}$$

$$\downarrow$$

A les pàgines de l'INE hem trobat aquestes dades:

$$IPC_{1992}^{1998} = 123{,}791 \qquad IPC_{2001}^{2001} = 100$$
$$IPC_{1992}^{2001} = 135{,}702 \qquad IPC_{2001}^{2006} = 117{,}624$$

I com que necessitaven esbrinar el valor de IPC_{2001}^{1998}, hem procedit com s'explica en l'exercici 4:

$$IPC_{2001}^{1998} = \frac{IPC_{1992}^{1998} \cdot IPC_{2001}^{2001}}{IPC_{1992}^{2001}} = \frac{123{,}791 \cdot 100}{135{,}702} = 91{,}223$$

Aquests 124961,01 euros serien el valor equivalent, quant a poder adquisitiu, del valor de compra de l'habitatge a l'any 2006.

Com que a l'enunciat es diu que l'hem venut per 240000 euros, calcularem els beneficis en termes relatius, a partir del concepte d'índex.

$$\frac{240000}{124961{,}01} = 1{,}92$$

Aquest quocient ens permet interpretar que tenim un benefici del 92%, és a dir, quasi s'ha duplicat el valor de l'habitatge en el període de 8 anys que hem contemplat.

Nota: Per a fer els càlculs hem utilitzat les mitjanes anuals de l'IPC, però es podria fer també amb els valors de l'IPC exactament dels mesos de compra i venda, així com triar els IPC del grup d'habitatge en lloc de l'IPC general. Deixem aquestes variants per al treball del lector.

Exercici 7

En la taula següent mostrem les dades dels impostos municipals de cert habitatge els darrers anys.

Any	*Import impost municipal (termes nominals)*
2006	503,24
2007	515,65
2008	536,73
2009	578,84
2010	584,42

Per a analitzar la seua evolució,

a) Deflacteu la sèrie, convertint-la en monedes constants de l'any 2006.
b) Calculeu els índexs que ens permetran estudiar la seua evolució any per any, en termes reals o monedes constants de l'any 2006. Interpreteu-ne els resultats.
c) Calculeu l'increment total i l'increment mitjà en el període en termes reals.
d) Suposant que els impostos seguisquen aquest comportament, esbrineu el valor en termes nominals o monedes corrents per als anys 2011, 2012 i 2013.

Nota: Per a resoldre aquest exercici, utilitzarem els valors de la mitjana anual de l'IPC general que necessitem, obtenint-los de la pàgina web de l'INE. www.ine.es

Solució

a) Deflacta la sèrie, convertint-la en termes reals o monedes constants de l'any 2006.

A tal fi, cal que consultem els valors de l'IPC d'aquests anys. Prendrem les mitjanes anuals de l'IPC general. Cal insistir en què tots els índexs que treballem al mateix exercici deuen estar en la mateixa base; en cas contrari, s'ha de fer el canvi de base que pertoque tal com es va explicar a l'exercici 4.

Presentem els resultats en la taula següent:

Any	Import impost municipal (termes nominals)	Import impost municipal (termes reals 2006)
2006	503,24	503,24
2007	515,65	$515{,}65 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2007}} = 515{,}65 \cdot \frac{100}{102{,}787} = 501{,}67$
2008	536,73	$536{,}73 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2008}} = 536{,}73 \cdot \frac{100}{106{,}976} = 501{,}73$
2009	578,84	$578{,}84 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2009}} = 578{,}84 \cdot \frac{100}{106{,}668} = 542{,}66$
2010	584,42	$584{,}42 \cdot \frac{IPC_{2006}^{2006}}{IPC_{2006}^{2010}} = 584{,}42 \cdot \frac{100}{108{,}588} = 538{,}20$

Amb aquesta operació, li hem «eliminat» a l'import de l'impost, l'efecte de la inflació i podrem analitzar «en termes reals» la seua evolució com a tal magnitud, llevat de les influències dels esdeveniments de l'economia general que es reflecteixen en les variacions de l'índex de preus.

b) Calculeu els índexs que ens permetran estudiar la seua evolució any per any, en termes reals o monedes constants de l'any 2006. Interpreteu-ne els resultats.

Ens demanen els índexs encadenats amb els valors de la darrera columna de la taula anterior:

Any	Import impost municipal (termes nominals)	Import impost municipal (termes reals 2006)	Índex	Interpretació
2006	503,24	503,24	-----	
2007	515,65	501,67	$I_{2006}^{2007} = \frac{501{,}67}{503{,}24} = 0{,}997$	Ha disminuït un 0,3%
2008	536,73	501,73	$I_{2007}^{2008} = \frac{501{,}73}{501{,}67} = 1{,}0001$	Podem considerar que és constant
2009	578,84	542,66	$I_{2008}^{2009} = \frac{542{,}66}{501{,}73} = 1{,}082$	Ha augmentat un 8,2%
2010	584,42	538,20	$I_{2009}^{2010} = \frac{538{,}20}{542{,}66} = 0{,}992$	Ha disminuït un 0,8%

En general, veiem que en termes reals és un import que roman estable en el període analitzat, ja que l'evolució de l'import és paral·lela a l'evolució de l'IPC, tret de l'any 2009 que, de manera puntual, fa un augment del 8,2%.

Podem veure amb més claredat aquesta evolució, quan hem «esborrat» l'efecte de la inflació.

c) Calculeu l'increment total i l'increment mitjà en el període en termes reals.

Per a calcular els increments que ens plantegem, és més còmode partir de les dades de la magnitud. En aquest cas, ens referim a l'import de l'impost municipal en termes reals de l'any 2006.

Per a determinar l'increment total del període, interpretarem l'índex:

$$I_{2006}^{2010} = \frac{538,20}{503,24} = 1,069$$

que ens permetrà afirmar que l'import de l'impost ha augmentat un 6,9% en termes reals en el període considerat.

Per a determinar l'increment mitjà anual, calcularem l'arrel següent:

$$\sqrt[4]{1,069} = 1,017$$

que ens permetrà afirmar que l'increment total del 6,9% és equivalent a un increment constant anual de l'1,7% durant 4 anys.

Volem assenyalar que aquests resultats també es podrien obtenir a partir dels índexs «encadenats», encara que no és raonable si disposem dels valors de la magnitud a analitzar.

Així, l'increment total del període seria:

$$0,997 \cdot 1,0001 \cdot 1,082 \cdot 0,992 = 1,07$$

que esdevindria un 7% d'augment total en el període. La diferència (una dècima) és deguda a les errades de l'arrodoniment de cadascun dels índexs.

L'increment mitjà anual s'obtindria també:

$$\sqrt[4]{0,997 \cdot 1,0001 \cdot 1,082 \cdot 0,992} = \sqrt[4]{1,0702} = 1,017$$

que dóna el mateix resultat que hem comentat abans.

d) Suposant que els impostos seguisquen aquest comportament, esbrineu el valor en termes nominals o monedes corrents per als anys 2011, 2012 i 2013.

Com que ens demanen que donem el resultat en moneda corrent o termes nominals haurem d'estimar els possibles valors de l'IPC en els anys venidors, per a operar de manera semblant a l'apartat *a*) però en sentit oposat.

A tal fi, obtindrem l'increment mitjà anual de l'IPC dels anys del període estudiat, a partir dels valors del primer i darrer any.

$\frac{IPC_{2006}^{2010}}{IPC_{2006}^{2006}} = \frac{108{,}588}{100} = 1{,}08588$ i per a obtenir l'increment mitjà anual de l'IPC del període calcularem l'arrel quarta corresponent:

$\sqrt[4]{1{,}08588} = 1{,}0208$ que ens permet afirmar que l'increment és del 2,08% anual. Així, basant-nos en aquest resultat, podrem estimar l'IPC dels anys següents:

$$IPC_{2006}^{2011} = IPC_{2006}^{2010} \cdot 1{,}0208 = 108{,}588 \cdot 1{,}0208 = 110{,}847$$
$$IPC_{2006}^{2012} = IPC_{2006}^{2010} \cdot 1{,}0208^2 = 108{,}588 \cdot 1{,}0208^2 = 113{,}152$$
$$IPC_{2006}^{2013} = IPC_{2006}^{2010} \cdot 1{,}0208^3 = 108{,}588 \cdot 1{,}0208^3 = 115{,}506$$

Estimarem primer l'import de l'impost en termes reals, aplicant l'increment mitjà anual (1,7%) sobre el valor d'aquest import de l'any 2010, és a dir, 538,20 euros en termes reals de l'any 2006.

Any 2011 → 538,20 · 1,017 = 547,25
Any 2012 → 538,20 · 1,017 ^{2} = 556,46
Any 2013 → 538,20 · 1,017 ^{3} = 565,82

Com que aquests resultats estan expressats en termes reals de l'any 2006, cal convertir-los en termes nominals:

$$\text{Any 2011} \rightarrow 547{,}25 \cdot \frac{IPC_{2006}^{2011}}{IPC_{2006}^{2006}} = 547{,}25 \cdot \frac{110{,}847}{100} = 606{,}61$$
$$\text{Any 2012} \rightarrow 556{,}46 \cdot \frac{IPC_{2006}^{2012}}{IPC_{2006}^{2006}} = 556{,}46 \cdot \frac{113{,}152}{100} = 629{,}65$$
$$\text{Any 2013} \rightarrow 565{,}82 \cdot \frac{IPC_{2006}^{2013}}{IPC_{2006}^{2006}} = 565{,}82 \cdot \frac{115{,}506}{100} = 653{,}56$$

Tan sols notar que aquesta estimació està feta sota la hipòtesi que tan l'IPC com l'import de l'impost, evolucionaràn al ritme anual que indique l'increment mitjà anual de cadascuna de les magnituds.

En aquest exercici, malgrat tot, sembla que l'increment no reflecteix la realitat del comportament de l'import de l'impost, el qual ha estat estable la major part del període i tan sols va experimentar un important augment del 8% l'any 2009, de manera puntual. Aquesta matisació fa que les estimacions, en certa manera, perden certa fiabilitat.

3k) Saber calcular les variacions del poder adquisitiu d'un salari, en funció de les variacions del salari i de l'IPC.

Exercici 8

En la taula següent indiquem el valor de la nòmina mensual d'un treballador els darrers anys.

Any	Nòmina mensual (€)
2007	2034,75
2008	2062,13
2009	2218,61
2010	2253,67
2011	2181,75

Estudia la pèrdua o guany del seu poder adquisitiu per a cada any i de tot el període global, considerant els valors de la mitjana anual de l'IPC general que pots trobar en la pàgina de l'INE.

Solució

Seria convenient recordar en aquest moment el concepte de *pèrdua o guany de poder adquisitiu*. Es pot dir que nosaltres guanyem poder adquisitiu (capacitat de compra de béns de consum) si el salari que percebem aquest any està per damunt del que percebríem si el nostre salari haguera estat incrementat en el mateix percentatge que augmenten els preus d'aquests béns. Podríem raonar de la mateixa manera per a definir la pèrdua de poder adquisitiu quan el nostre salari queda per davall del que tindríem si l'hagueren incrementat amb el mateix percentatge que els preus. El seu increment està reflectit en l'IPC que publica l'INE cada mes. Nosaltres agafarem la mitjana anual general d'aquest índex que podrem trobar fàcilment en la web d'aquest organisme.

Ara bé, com fem una anàlisi en termes relatius i donem el resultat en percentatges, vegem en la següent expressió, com el salari concret del qual partim, no és necessari en l'estudi de l'evolució del poder adquisitiu:

Guany o pèrdua de poder adquisitiu = $\Delta_{poder\ adquisitiu} =$

$$= \frac{salari\ \ nou}{salari\ \ actualitzat\ \ segons\ \ IPC} = \frac{salari\ \ anterior \cdot (1+\Delta_{salarial})}{salari\ \ anterior \cdot (1+\Delta_{IPC})} = \frac{(1+\Delta_{salarial})}{(1+\Delta_{IPC})}$$

La pèrdua o guany del poder adquisitiu, doncs, es calcula a partir de la comparació dels increments anuals del salari (Δ_{salari}) i de l'IPC (Δ_{IPC},) paral·lelament.

A tal fi, començarem per calcular els índexs «encadenats» dels salaris, que ens permetran esbrinar els increments salarials anuals. A la següent taula es detallen els càlculs i els resultats.

Any	*Nòmina mensual (€)*	Índex	
2007	2034,75	------	-------
2008	2062,13	$I_{2007}^{2008} = \frac{2062,13}{2034,75} = 1,014$	Ha augmentat 1,4%
2009	2218,61	$I_{2008}^{2009} = \frac{2218,61}{2062,13} = 1,076$	Ha augmentat 7,6%
2010	2253,67	$I_{2009}^{2010} = \frac{2253,67}{2218,61} = 1,016$	Ha augmentat 1,6%
2011	2181,75	$I_{2010}^{2011} = \frac{2181,75}{2253,67} = 0,968$	Ha disminuït 3,2%

Consultarem la pàgina de l'INE per a trobar els valors de l'IPC d'aquests anys. Ens interessa la mitjana anual de l'índex general en base 2006. Com que l'any 2011 no està encara publicat en el moment de l'elaboració d'aquests materials, utilitzarem l'estimació que vam obtenir a l'exercici 7 d'aquesta col·lecció. I amb aquestes dades farem la mateixa anàlisi que hem fet amb les nòmines per a obtenir els increments anuals.

Any	*IPC* *Base 2006*	Índex	
2007	$IPC_{2006}^{2007} = 102,787$	------	-------
2008	$IPC_{2006}^{2008} = 106,976$	$I_{2007}^{2008} = \frac{106,976}{102,787} = 1,041$	Ha augmentat 4,1%
2009	$IPC_{2006}^{2009} = 106,668$	$I_{2008}^{2009} = \frac{106,668}{106,976} = 0,997$	Ha disminuït 0,3%
2010	$IPC_{2006}^{2010} = 108,588$	$I_{2009}^{2010} = \frac{108,588}{106,668} = 1,018$	Ha augmentat 1,8 %
2011	$IPC_{2006}^{2011} = 110,847$	$I_{2010}^{2011} = \frac{110,847}{108,588} = 1,021$	Ha augmentat 2,1%

Ja hem explicat en començar l'exercici el concepte de *pèrdua o guany de poder adquisitiu*. Hom guanya poder adquisitiu quan l'increment salarial està per damunt de l'increment de l'IPC que ens indica, així mateix, l'increment dels preus dels béns de consum. De la mateixa manera, hi haurà una pèrdua de poder adquisitiu quan l'increment salarial estiga per sota de l'increment de l'IPC.

Per a fer aquesta comparació, partirem dels índexs que hem calculat en les dues taules anteriors i com que es fa un estudi en termes relatius farem els quocients d'aquestes quantitats any per any. Mostrem els resultats en la taula següent i els càlculs de les caselles ombrejades en gris, davall de la taula.

	2008	2009	2010	2011
Increment salarial	1,014	1,076	1,016	0,968
Increment IPC	1,041	0,997	1,018	1,021
Pèrdua o guany poder adquisitiu	0,974	1,079	0,998	0,948

Any 2008 $\rightarrow \dfrac{1{,}014}{1{,}041} = 0{,}974$ Any 2009 $\rightarrow \dfrac{1{,}076}{0{,}997} = 1{,}079$

Any 2010 $\rightarrow \dfrac{1{,}016}{1{,}018} = 0{,}998$ Any 2011 $\rightarrow \dfrac{0{,}968}{1{,}021} = 0{,}948$

Queda més clar si anotem les interpretacions en percentatges, i així ho mostrem a la taula següent. Convindrem que el signe positiu indica guany de poder adquisitiu i el negatiu n'indica pèrdua.

	2008	2009	2010	2011
Increment salarial	+1,4%	+7,6%	+1,6%	-3,2%
Increment IPC	+4,1%	-0,3%	+1,8%	+2,1%
Pèrdua o guany poder adquisitiu	-2,6%	+7,9%	-0,2%	-5,2%

Si volem analitzar l'increment total dels tres conceptes, podem utilitzar les magnituds originals de què disposem tant pel que respecta als salaris com a l'IPC, i ho farem a partir dels seus índexs, per al poder adquisitiu.

Vegem l'increment salarial del total del període:

$$I_{salaris}{}_{2007}^{2011} = \frac{2181{,}75}{2034{,}75} = 1{,}072 \rightarrow \text{Ha augmentat un } 7{,}2\%$$

Calculem ara l'increment de l'IPC (recordem que la data de l'IPC del 2011 és una estimació).

$$I_{IPC\,2007}^{\;\;2011} = \frac{IPC_{2006}^{2011}}{IPC_{2006}^{2007}} = \frac{110{,}847}{102{,}787} = 1{,}078 \rightarrow \text{Ha augmentat un } 7{,}8\%$$

Per a calcular la pèrdua de poder adquisitiu, plantegem el quocient d'aquests increments en la seua expressió d'índex.

$$I_{poder\ adquisitiu\,2007}^{\;\;2011} = \frac{I_{salaris\,2007}^{\;\;2011}}{I_{IPC\,2007}^{\;\;2011}} = \frac{1{,}072}{1{,}078} = 0{,}994 \rightarrow \text{Ha disminuït un } 0{,}6\%$$

Si completem la taula anterior amb aquesta informació, tenim detallada l'evolució total.

	2008	2009	2010	2011	TOTAL
Increment salarial	+1,4%	+7,6%	+1,6%	-3,2%	+7,2%
Increment IPC	+4,1%	-0,3%	+1,8%	+2,1%	+7,8%
Pèrdua o guany poder adquisitiu	-2,6%	+7,9%	-0,2%	-5,2%	-0,6%

Advertim que encara que amb una ullada ens puga parèixer que els resultats de les caselles ombrejades es podrien obtenir sumant i restant els percentatges en files i columnes, cal fixar-se que no és cert tal com es pot comprovar amb les dades totals i en algunes columnes, per exemple l'any 2008.

Ara bé, sí podem obtenir els resultats, a partir dels índexs de la taula prèvia, multiplicant-los.

Vegem l'increment salarial del total del període:

$$I_{salaris\,2007}^{\;\;2011} = 1{,}014 \cdot 1{,}076 \cdot 1{,}016 \cdot 0{,}968 = 1{,}073 \rightarrow \text{Ha augmentat un } 7{,}3\%$$

Calculem ara l'increment de l'IPC.

$$I_{IPC\,2007}^{\;\;2011} = 1{,}041 \cdot 0{,}997 \cdot 1{,}018 \cdot 1{,}021 = 1{,}079 \rightarrow \text{Ha augmentat un } 7{,}9\%$$

Per a calcular la pèrdua de poder adquisitiu, plantegem també el producte dels factors:

$$I_{poder\ adquisitiu\,2007}^{\;\;2011} = 0{,}974 \cdot 1{,}079 \cdot 0{,}998 \cdot 0{,}948 = 0{,}994 \rightarrow \text{Ha disminuït un } 0{,}6\%$$

Cal advertir que la diferència amb els resultats anteriors (de l'ordre de dècimes), és deguda a l'arrodoniment de cada factor. Per aquesta raó insistim en la recomanació d'utilitzar les dades originals de les magnituds a analitzar si en disposem, però presentem els dos mètodes de resolució, per als casos en què la informació disponible siguen els increments percentuals anuals.

Exercici 9

En les taules següents es presenten els valors de l'IPC i l'increment salarial d'un treballador en els anys que indiquem en certa comunitat.

Anys	IPC
2008	115,1
2009	119,2
2010	121,6
2011	123,8

Anys	Increment salarial anual (%) Anual IPC
2008	
2009	1,8
2010	2,7
2011	1,7

a) Calculeu l'increment mitjà i total del salari en el període 2008-2011.
b) Calculeu l'increment anual, mitjà i total de l'IPC en el període 2008-2011.
c) Si suposem que les condicions econòmiques de la comunitat no varien, realitzeu una previsió del valor de l'IPC per a l'any 2013.
d) Estudieu per a cada any i per al període total la pèrdua o guany del poder adquisitiu i realitzeu una interpretació de les dades obtingudes.

Solució

Seria convenient recordar en aquest moment el concepte de *pèrdua o guany de poder adquisitiu* que pots trobar a l'exercici núm. 8.

$$\textit{Guany o pèrdua de poder adquisitiu} = \Delta_{poder\ adquisitiu} =$$

$$= \frac{salari\ \ nou}{salari\ \ actualitzat\ \ segons\ \ IPC} = \frac{salari\ \ anterior \cdot (1+\Delta_{salarial})}{salari\ \ anterior \cdot (1+\Delta_{IPC})} = \frac{(1+\Delta_{salarial})}{(1+\Delta_{IPC})}$$

La pèrdua o guany del poder adquisitiu es calcula a partir de la comparació dels increments anuals del salari (Δ_{salari}) i de l'IPC (Δ_{IPC},) paral·lelament.

A tal fi, com que a l'enunciat ja disposem dels increments anuals dels salaris, caldrà calcular els increments mitjà i total en el període 2008-2011.

Any	Salaris	Índex
2008		
2009	Ha augmentat 1,8%	$I_{2008}^{2009}=1{,}018$
2010	Ha augmentat 2,7%	$I_{2009}^{2010}=1{,}027$
2011	Ha augmentat 1,7%	$I_{2010}^{2011}=1{,}017$

a) Per a calcular l'increment mitjà del salaris del període considerarem que tenim tres índexs, i així calcularem l'arrel tercera del producte d'aquests factors:

$\sqrt[3]{1{,}018 \cdot 1{,}027 \cdot 1{,}017}=\sqrt[3]{1{,}0633}=1{,}0207$ que interpretem com que els salaris han augmentat una mitjana del 2,07% anual.

Per a calcular l'increment total del període ho farem a partir dels índexs encadenats:

$I_{2008}^{2011}=I_{2010}^{2011} \cdot I_{2009}^{2010} \cdot I_{2008}^{2009}=1{,}017 \cdot 1{,}027 \cdot 1{,}018=1{,}0633$ que interpretem com que els salaris han augmentat un 6,33 % al llarg dels tres anys.

b) Per a calcular l'increment anual, mitjà i total de l'IPC en el període 2008-2011, determinarem la seqüència dels índexs encadenats que presentem a continuació, on podem veure els increments anuals:

Any	IPC	Índex	
2008	115,1		
2009	119,2	$I_{2008}^{2009}=\frac{119{,}2}{115{,}1}=1{,}0356$	Ha augmentat 3,56 %
2010	121,6	$I_{2009}^{2010}=\frac{121{,}6}{119{,}2}=1{,}0201$	Ha augmentat 2,01 %
2011	123,8	$I_{2010}^{2011}=\frac{123{,}8}{121{,}6}=1{,}0181$	Ha augmentat 1,81%

Per a calcular l'increment total del període podem operar a partir dels índexs encadenats:

$$I_{2008}^{2011}=I_{2010}^{2011} \cdot I_{2009}^{2010} \cdot I_{2008}^{2009}=1{,}0181 \cdot 1{,}0201 \cdot 1{,}0356=1{,}0756$$

o també a partir dels valors de l'IPC inicials i finals del període total:

$$I_{2008}^{2011}=\frac{123{,}8}{115{,}1}=1{,}0756$$

interpretarem que l'IPC ha augmentat un 7,56% en tot el període.

Per a calcular l'increment mitjà de l'IPC, també podem operar paral·lelament:

$$\sqrt[3]{I_{2010}^{2011} \cdot I_{2009}^{2010} \cdot I_{2008}^{2009}} = \sqrt[3]{1{,}0181 \cdot 1{,}0201 \cdot 1{,}0356} = \sqrt[3]{1{,}0756} = 1{,}0246$$

o obrar així:

$$\sqrt[3]{I_{2008}^{2011}} = \sqrt[3]{\frac{123{,}8}{115{,}1}} = \sqrt[3]{1{,}0756} = 1{,}0246$$

interpretarem que l'IPC ha augmentat una mitjana del 2,46% cada any.

c) Si les condicions econòmiques de la comunitat suposem que no varien, per a realitzar una previsió del valor de l'IPC per a l'any 2013, partirem de la darrera dada coneguda de l'IPC de l'any 2011 i l'incrementarem amb el percentatge que hem obtingut com a increment mitjà en l'apartat anterior:

$$IPC_{2013} = IPC_{2011} \cdot (Inc\ mitjà\ anual)^2 = 123{,}8 \cdot 1{,}0246^2 = 129{,}97$$

Per a calcular les variacions del poder adquisitiu presentarem les dades dels increments en la taula següent:

Any	2009	2010	2011	Total
Inc. salarial	1,8	2,7	1,7	6,33
Inc. IPC	3,56	2,01	1,81	7,56
Inc. poder adquisitiu	–1,7	+0,68	-0,11	-1,14

Ja hem explicat, al començament de l'exercici el concepte de *pèrdua o guany de poder adquisitiu*. Hom guanya poder adquisitiu quan l'increment salarial està per damunt de l'increment de l'IPC que ens indica, així mateix, l'increment dels preus dels béns de consum. De la mateixa manera, hi haurà una pèrdua de poder adquisitiu quan l'increment salarial estarà per sota de l'increment de l'IPC.

Per a fer aquesta comparació, partirem dels índexs que hem calculat en les dues taules anteriors, i com que es fa un estudi en termes relatius, farem els quocients d'aquestes quantitats any per any. Mostrem els resultats en la taula anterior i els càlculs de les caselles ombrejades en gris a continuació:

$$\text{Any 2009} \rightarrow \frac{1{,}018}{1{,}0356} = 0{,}9830 \rightarrow -1{,}7\%$$

$$\text{Any 2010} \rightarrow \frac{1{,}027}{1{,}0201} = 1{,}0068 \rightarrow +0{,}68\%$$

$$\text{Any 2011} \rightarrow \frac{1{,}017}{1{,}0181} = 0{,}9989 \rightarrow -0{,}11\%$$

Per a estudiar el període total → $\frac{1{,}0633}{1{,}0756} = 0{,}9886 \rightarrow -1{,}14\%$

Queda més clar si anotem les interpretacions en percentatges, i així ho mostrem a la taula anterior. Convindrem que el signe positiu indica guany de poder adquisitiu i el negatiu n'indica pèrdua.

Exercici 10

Per a fer un estudi de l'evolució del preu de cert model d'ordinador en termes reals, disposem de les dades que presentem a la taula següent:

a) Calculeu l'increment anual, mitjà i total del preu de l'ordinador en termes reals.
b) Si seguim aquesta evolució, estimeu el preu que podria tenir l'ordinador l'any 2008.

c)

	2000	2001	2002	2003	2004
IPC base 1992	131	135	139		
IPC base 2002			103	106	109
	1300	1275	1250	1100	950

Nota: hem de recórrer a períodes i valors molt antics o imaginats per a treballar l'objectiu del canvi de base de l'IPC, ja que amb la nova metodologia del càlcul de l'IPC per l'INE aquesta circumstància s'ha superat, pero és important que l'alumnat conega aquest contingut per a no treballar en sèries d'IPC no adients en un mateix exercici.

Solució

Per començar, cal fer un canvi de base i trobar tots els valors de l'IPC en la mateixa base. Emplenarem la taula amb els nous valors trobats (en roig) i davall anotarem els càlculs realitzats.

	2000	2001	2002	2003	2004
IPC base 1992	131	135	139	143,05	147,10
IPC base 2002	97,07	100,04	103	106	109

Vegem les dades de l'any 2003:

$$IPC_{2002}^{2002} = 103 \rightarrow IPC_{2002}^{2003} = 106$$

$$IPC_{1992}^{2002} = 139 \quad \rightarrow \quad IPC_{1992}^{2003} = X$$

Per a calcular la dada desconeguda:

$$IPC_{1992}^{2003} = \frac{IPC_{2002}^{2003} \cdot IPC_{1992}^{2002}}{IPC_{2002}^{2002}} = \frac{109 \cdot 139}{103} = 106\frac{139}{103} = 106 \cdot 1{,}35 = 143{,}05$$

De la mateixa manera, calculem la dada de l'any 2004:

$$IPC_{2002}^{2002} = 103 \quad \rightarrow \quad IPC_{2002}^{2004} = 109$$
$$IPC_{1992}^{2002} = 139 \quad \rightarrow \quad IPC_{1992}^{2004} = X$$

Per a calcular la dada desconeguda:

$$IPC_{1992}^{2004} = \frac{IPC_{2002}^{2004} \cdot IPC_{1992}^{2002}}{IPC_{2002}^{2002}} = \frac{109 \cdot 139}{103} = 106\frac{139}{103} = 109 \cdot 1{,}35 = 147{,}10$$

De la mateixa manera podríem plantejar (per proporcionalitat) els càlculs per a obtenir la resta de dades de la base 2002.

Per a continuar l'exercici utilitzarem els IPC en aquesta base 1992 o en 2002. És indiferent sempre que anem amb compte de treballar tots els índexs en la mateixa base.

a) Calculeu l'increment anual, mitjà i total del preu de l'ordinador en termes reals.

Com que l'anàlisi del preu de l'ordinador ens la demanen en termes reals, cal fer primer la conversió del preu de l'ordinador a moneda constant de l'any 2000 (deflactació de la sèrie de preus de l'ordinador).

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)
2000	1300	1300
2001	1275	$1275 \cdot \frac{IPC_{1992}^{2000}}{IPC_{1992}^{2001}} = 1275 \cdot \frac{131}{135} = 1237{,}2$
2002	1250	$1250 \cdot \frac{IPC_{1992}^{2000}}{IPC_{1992}^{2002}} = 1250 \cdot \frac{131}{139} = 1178{,}06$
2003	1100	$1100 \cdot \frac{IPC_{1992}^{2000}}{IPC_{1992}^{2003}} = 1100 \cdot \frac{131}{143{,}05} = 1007{,}34$
2004	950	$950 \cdot \frac{IPC_{1992}^{2000}}{IPC_{1992}^{2004}} = 950 \cdot \frac{131}{147{,}10} = 846{,}02$

Per a calcular l'increment anual del preu de l'ordinador en termes reals, operem en la taula següent mitjançant índexs encadenats:

Any	Preu ordinador (termes nominals)	Preu ordinador (termes reals 2000)	Índex	Interpretació: Increment anual
2000	1300	1300	-----	
2001	1275	1237,2	$I_{2000}^{2001} = \frac{1237,2}{1300} = 0,9517$	Ha disminuït un 4,83%
2002	1250	1178,06	$I_{2001}^{2002} = \frac{1178,06}{1237,2} = 0,9522$	Ha disminuït un 4,78%
2003	1100	1007,34	$I_{2002}^{2003} = \frac{1007,34}{1178,06} = 0,8551$	Ha disminuït un 14,49%
2004	950	846,02	$I_{2003}^{2004} = \frac{846,02}{1007,34} = 0,8399$	Ha disminuït un 16,01%

Per a calcular l'increment total i mitjà del període, operem:

$I_{2000}^{2004} = \frac{846,02}{1300} = 0,6508$ $\rightarrow$ al llarg del període ha disminuït un 34,92% el seu valor en termes reals, que equival a un increment mitjà anual de:

$\sqrt[4]{\frac{846,02}{1300}} = \sqrt[4]{0,6508} = 0,8982$ $\rightarrow$ una disminució anual mitjana del 10,18%

b) Si seguim aquesta evolució, estimem el preu que podria tenir l'ordinador l'any 2008.

Per a fer aquest apartat, considerem que ens demanen l'estimació en termes corrents del preu, suposant que no varia el comportament de l'IPC ni l'evolució del preu de l'ordinador en termes reals que hem analitzat a l'apartat anterior.

Per a fer aquestes estimacions necessitem l'increment mitjà de les dues sèries. Ens falta calcular l'increment mitjà de l'IPC en el període que ens ocupa:

L'IPC ha augmentat un 2,94% anualment, per la qual cosa podem estimar el valor de l'IPC a l'any 2998.

$$IPC_{1992}^{2008} = IPC_{1992}^{2004} \cdot 1,0294^4 = 147,10 \cdot 1,0294^4 = 165,18$$

Farem les mateixes operacions amb la sèrie del preus dels ordinadors en termes reals.

Preu ordinador per a l'any 2008 en termes reals del 2000 = 846,02 · $0,8982^4$ = 550,65 €

Per a passar-los a termes corrents en moneda del 2008,

$$550{,}65 \cdot \frac{IPC_{1992}^{2008}}{IPC_{1992}^{2000}} = 550{,}65 \cdot \frac{165{,}18}{131} = 694{,}3 \text{ €}$$

UNITAT 4

Sèries temporals

Introducció teòrica

Com a elements introductoris d'aquest capítol, és convenient recordar definicions de conceptes que necessitarem per a assolir els objectius a treballar en aquesta unitat (referències bibliogràfiques 1, 21 i 24).

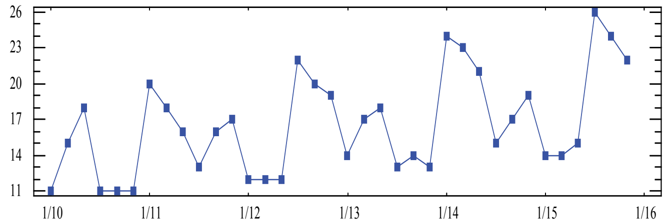
Sèrie temporal

És una successió d'observacions quantitatives d'un fenomen ordenades en els temps i períodes equidistants. Cada observació que denotarem per y_{ij}, correspon al valor de la magnitud en l'any i i període j. Exemple, si tenim una sèrie d'observacions trimestrals al llarg dels anys 2001 al 2005, entendrem que la dada x_{23} correspon al valor corresponent al 3r trimestre de l'any 2002.

Gràfica d'una sèrie temporal

La primera eina descriptiva que ens pot permetre analitzar una sèrie es la seua gràfica, que dibuixarem situant els valors de la sèrie a l'eix d'ordenades i els valors dels períodes a l'eix d'abscisses.

Detallarem més endavant la importància d'observar aquest gràfic per a assegurar-nos d'ajustar la nostra sèrie a un model additiu i per a confirmar els resultats de les components, que han de reflectir quantitativament uns valors que confirmen la nostra visió del fenomen a la gràfica.



Components d'una sèrie temporal

En l'anàlisi d'una sèrie temporal en aquest tema considerarem que tota sèrie empírica analitzada estarà formada per quatre components teòriques: *tendència*, *variacions cícliques, variacions estacionals* i *variacions residuals*.

Tendència: és la component que ens explica el comportament del fenomen a «llarg termini». La denotarem per T_{ik} i ens permetrà explicar si les mitjanes anuals dels valors de la sèrie augmenten o disminueixen en el període que volem analitzar.

Variacions cícliques: són variacions que es produeixen amb una periodicitat superior a l'any i freqüentment es manifesten com a conseqüència de períodes de prosperitat i de depressió en l'activitat econòmica, o en altres magnituds qualssevol. Les denotem per c_{ik}. No les obtindrem en aquest tema, ja que queden fora de l'abast de les tècniques que desenvoluparem per a taules menudes a fi d'estudiar el desenvolupament i justificació teòrics i explícit dels càlculs. Per a taules més grans ens ajudarem de programari que ens donarà resultats que caldrà analitzar amb les consideracions teòriques que podrem veure en els exercicis proposats amb taules de menor tamany.

Variacions estacionals: són oscil·lacions que es produeixen amb una periodicitat dintre de l'any i que es poden identificar repetidament al llarg dels anys dels quals disposem dades per analitzar. Per exemple, històricament les sèries de l'atur augmenten a l'hivern i disminueixen a l'estiu, el volum de vendes d'una superfície comercial té pujades significatives en períodes de rebaixes, etc. Les podrem mesurar en valors absoluts (component estacional e_k) o en valors relatius (índexs estacionals I_k) respecte a la mitjana global (M = mitjana aritmètica de les mitjanes corregides).

Variacions residuals o erràtiques: ja que les dades son empíriques, és d'esperar que, de manera natural, n'hi haja lleus variacions aleatòries respecte al model teòric que pretén analitzar la sèrie amb la informació de la resta dels components. Les denotarem per $\boldsymbol{r}_{ik}$ i també s'anomenaran residus. Cal que no presenten periodicitat manifesta i siguen de valor reduït. Quan qualsevol dels seus valors ens cride l'atenció pel seu valor absolut respecte a la resta, ens indicarà una dada que per qualsevol motiu no s'ajusta al model que pretenem obtenir. Cal analitzar el seu origen: errada, efecte produït per una vaga, un accident, una pertorbació meteorològica, que caldrà trobar amb la informació pertinent a l'abast del context de la sèrie estudiada i que intentarem explicar per a justificar la variació d'aquestes dades en particular que es desajusten del model.

Nosaltres estudiarem tan sols sèries temporals que suposarem que s'ajusten al *model additiu*, circumstància que es pot comprovar acudint a la bibliografia que referim i que no hem desenvolupat en aquesta col·lecció de problemes. Per això podem considerar que una dada en particular és el resultat de la suma de les seues components. Així:

$$y_{ij} = T_{ik} + c_{ik} + e_{ik} + r_{ik}$$

Per a fer l'anàlisi d'una sèrie estudiarem dos mètodes: ajust analític i el mètode de les mitjanes mòbils. A continuació posarem el formulari i notació de cadascun dels mètodes que es podran seguir en els exercicis resolts a continuació:

Mètode de l'ajust analític

Per treballar aquest mètode, és convenient presentar els càlculs en forma de taula (veure exemples als exercicis resolts) convenientment ordenats.

Per a determinar la recta de tendència, a la part inferior de la taula de les dades, calcularem per columnes les mitjanes mensuals de cada any $\overline{y}_i$, els valors de l'escala i, i a les dues files inferiors $\overline{y}_i \cdot i$, i^2, que ens permetrà calcular en la columna *totals* la suma dels valors de cada fila.

Cal explicar que la fila i, és una escala que crearem per a facilitar la resolució del sistema lineal de dues equacions amb dues incògnites. El procediment ens donarà resultats òptims situant el valor de l'any 0 en la columna central de la taula en el cas de tenir un nombre senar d'anys a la taula, o a qualsevol de les columnes adjacents en el cas d'un nombre parell d'anys a estudiar:

Exemple:

						TOTALS
Any	2008	2009	2010	2011	2012	
i	-2	-1	0	1	2	0

							TOTALS
Any	2008	2009	2010	2011	2012	2013	
i	-2	-1	0	1	2	3	3

Amb aquestes dades, que hem acumulat a la columna de *totals* plantejarem i resoldrem el sistema següent:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases}$$

on els coeficients a i b són els coeficients de la recta de regressió que denominem recta de tendència, la formula de la qual és $T_i = a + b$, en la qual i fa referència a l'any que s'indica en l'escala de les taules superiors dels exemples i N és el nombre d'anys o columnes que té la taula de les dades.

Aquesta recta que trobem no es sinó la recta d'ajust lineal per mínims quadrats a les mitjanes anuals $\overline{y}_i$.

Per la seua interpretació ens fixarem en el signe de la pendent (coeficient b) que ens determinarà un fenomen creixent o decreixent, segons el signe de b siga positiu o negatiu, respectivament i el valor de l'increment mitjà anual de la mitjana anual dels valors de la sèrie.

El valor de la tendència el denotarem per T_i i el considerarem constant per a totes les dades de l'any i.

Per a calcular la component estacional, treballarem les columnes que es poden veure a la dreta de la taula original.

A la primera columna, podem trobar les mitjanes aritmètiques dels valors originals de les dades de cada període o fila $\bar{y}_k = \frac{\sum_i y_{ij}}{N}$.

La columna següent correspon a les mitjanes corregides, $\bar{y}_k'$, la fórmula de les quals és:

$$\bar{y}_k' = \bar{y}_k - \frac{b}{m}(k-1)$$

on *b/m* podem interpretar-lo com l'increment que correspondria a cada període de l'increment anual de les dades, per la tendència del fenomen. Per això es corregeix aquest increment amb la fórmula abans indicada.

Notem que m és el nombre de files de la taula original, que correspon al nombre d'observacions que disposem en cada any. Així, $m = 12$ si es tracta d'observacions mensuals, $m = 4$ si es tracta d'observacions trimestrals, etc.

M és la mitjana global corregida i és la mitjana de les mitjanes corregides abans definides:

$$M = \frac{\sum_k \bar{y}_k'}{m}$$

que podem interpretar com el valor mitjà de les noves mitjanes corregides i que representarà el 100% o valor de referència, front a la que es comparen els comportaments estacionals que calculem en les dues columnes de la dreta de la taula i que són la component estacional $e_k = \bar{y}_k' - M$ i els índexs estacionals $I_k = \frac{\bar{y}_k'}{M} \cdot 100$. Amb aquests resultats podrem interpretar en quins períodes els valors de les observacions estan per damunt o per sota el valor de M en quantitats absolutes e_k ens presenta aquesta desviació mentre que I_k ho indica de manera percentual.

Per a calcular la component residual r_{ik} caldrà determinar primerament el valor de la tendència T_i per a cada any considerat en la taula, substituint en la recta de tendència el valor de i corresponent i considerarem per a la component estacional e_k els valors que ja hem calculat i explicat en els paràgrafs anteriors.

Així, per a cada observació, operarem $r_{ik} = y_{ik} - T_i - e_k$ i disposarem els resultats en la distribució de la taula original per a facilitar la seua interpretació i identificació del període i any corresponents.

Fem aquest advertiment perquè tots sabem que les quantitats que cal obtenir de la component residual haurien de ser menudes en valor absolut, i que no presenten cap regularitat. Ja sabem que estem calculant les quantitats no explicades pel nostre model i que permetran ressaltar aquells valors puntuals, que per raons no predictibles, mostren divergència del valor que caldria esperar, atenent les components de la tendència i estacional.

Recordem que fer prediccions per als anys propers implica que l'anàlisi del nostre model siga vigent i que cap altra circumstància aliena altere les regularitats que hem ressenyat amb el nostre model (la tendència explicada i el comportament dels períodes ja quantificat).

Si volem preveure les quantitats dels propers anys caldrà calcular els valors de la seua tendència T_i substituint en l'equació de la seua recta els valors de i que els correspondria en el cas que la taula continuara.

També considerarem els valors obtinguts de la component estacional e_k abans esmentada, i podrem fer les previsions de les dades futures, operant $y_{ik} = T_i + e_k$.

Mètode de les mitjanes mòbils

Aquest mètode està basat en el «suavitzat» d'una sèrie quan aquesta es substituïda per una successió de mitjanes aritmètiques de p observacions com explicarem a continuació. En aquest apartat teòric (càlcul de les mitjanes mòbils) canviarem la notació de la sèrie d'observacions inicial i passarem a ordenar-la amb un únic subíndex, considerant-la com una successió ordenada sense contemplar la seua procedència de període i any.

Per a aplicar aquest mètode començarem per triar un nombre p d'observacions a amitjanar amb uns criteris que després explicarem.

Si p es senar, formarem una sèrie nova de mitjanes que serà:

$$\overline{y}_{\frac{p+1}{2}} = \frac{y_1 + y_2 + \ldots + y_p}{p},\ \overline{y}_{\frac{p+3}{2}} = \frac{y_2 + y_3 \ldots + y_p + y_{p+1}}{p},\ \overline{y}_{\frac{p+5}{2}} = \frac{y_3 + y_4 \ldots + y_{p+1} + y_{p+2}}{p}$$

pot veure's que els subíndexs que adjudiquem a aquestes mitjanes obtingudes $\frac{p+1}{2}$, $\frac{p+3}{2}$, $\frac{p+5}{2}$ corresponen a un nombre enter, per la qual cosa aquestes mitjanes podem fer-les correspondre a un període original, ja que correspon al centre dels períodes amitjanats.

Si p es parell, aquesta circumstància no es dóna, ja que $\frac{p+1}{2}$, $\frac{p+3}{2}$, $\frac{p+5}{2}$ no corresponen en aquest cas a un nombre enter, per la qual cosa, davant la impossibilitat de fer correspondre les mitjanes aritmètiques a cap període de la sèrie original, realitzarem els càlculs de la mateixa manera i posteriorment farem un «centrat» determinant la mitjana aritmètica de cada dues mitjanes mòbils consecutives abans calculades.

Així:

$\overset{=}{y}_{\frac{p+2}{2}} = \dfrac{\overline{y}_{\frac{p+1}{2}} + \overline{y}_{\frac{p+3}{2}}}{2}$, $\overset{=}{y}_{\frac{p+4}{2}} = \dfrac{\overline{y}_{\frac{p+3}{2}} + \overline{y}_{\frac{p+5}{2}}}{2}$, $\overset{=}{y}_{\frac{p+6}{2}} = \dfrac{\overline{y}_{\frac{p+5}{2}} + \overline{y}_{\frac{p+7}{2}}}{2}$... ja que ara els nous subíndexs $\dfrac{p+2}{2}$, $\dfrac{p+4}{2}$, $\dfrac{p+6}{2}$... sí es corresponen a nombres enters i conseqüentment a períodes concrets de la sèrie inicial d'observacions.

Aquest mètode es basa en què la nova sèrie de mitjanes mòbils ens permetrà entreveure la tendència de la sèrie original a «llarg termini» ja que se suavitza el valor individual de cadascuna de les dades i les oscil·lacions. Perquè aquesta afirmació siga certa cal triar convenientment el nombre p d'observacions a amitjanar com hem indicat abans i anem a detallar a continuació.

El nombre p cal que siga múltiple del nombre d'observacions anuals (m) a fi de considerar en cada mitjana totes les fluctuacions estacionals. Així, a causa del mètode de construcció de les mitjanes mòbils abans esmentat, en cada mitjana se substituirà una dada que correspon a un cert període per una altra que es correspon al mateix període en la mitjana següent, i sempre tenim assegurada en cada càlcul la mitjana de totes les oscil·lacions dels diferents períodes dintre d'un any o més.

L'altre criteri a considerar es basa en l'observació de la gràfica de la sèrie original, la importància de la qual ja hem comentat en començar aquest apartat teòric. Observant aquesta gràfica cal trobar una certa periodicitat superior a l'any, és a dir, s'ha d'anular l'efecte d'una component estacional, agafant un nombre p que ha de ser múltiple del nombre d'observacions q que comprén el «període» gràfic que es repeteix al llarg de la sèrie.

Per tant, per a calcular la tendència, caldrà agafar un nombre p d'observacions que siga el mínim comú múltiple de m i q. Per exemple, si tenim una sèrie de dades trimestrals, m = 4 i de l'observació de la gràfica podem veure un patró bianual que es repeteix aproximadament cada 8 observacions, en aquest cas el nombre p a considerar per al càlcul de les mitjanes mòbils cal que siga 8, i per tractar-se d'un nombre parell serà necessari fer després un posterior «centrat».

Per a calcular la component estacional, determinarem per una banda la mitjana aritmètica de totes les dades que corresponen a cada període i que denotarem per $\overline{y_k}$, per altra banda, calculem les mitjanes mòbils amb $p = m$, nombre d'observacions anuals. En el cas que la sèrie tinga una component cíclica anual, podrem aprofitar els càlculs de l'apartat de la tendència.

Tot seguit disposarem les mitjanes mòbils en la distribució bidimensional que originàriament tenien les dades, distribuint-les en anys per columnes i en períodes per files. Podrem observar que el mètode de les mitjanes mòbils centrades obliga a que algunes de les cel·les de la taula queden buides, ja que no podem fer correspondre cap dada als períodes inicials i finals.

A continuació calcularem la mitjana aritmètica de les mitjanes que corresponen a cada període i que denotarem per E_k, ja que podem considerar-la com la component extraestacional, perquè el comportament estacional ha estat anul·lat per la tria de les dades convenient per a fer les mitjanes mòbils amb $p = m$.

Així: $\overline{e_k} = \overline{y_k} - \overline{E_k}$.

Per finalitzar aquesta part, recordem que els temes següents seran el desenvolupament de la probabilitat i la inferència (referències bibliogràfiques 7 i 26).

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

4*a*) Reconèixer en una col·lecció de dades els patrons i notació d'una sèrie temporal.
4*b*) Analitzar una sèrie i a partir de la seua gràfica, poder comprovar que s'adapta al model additiu.
4*c*) Conèixer les diferents components d'una sèrie temporal: tendència, component estacional, component erràtica o residual.
4*d*) Saber calcular les diferents components d'una sèrie, suposant un model additiu, pel mètode de l'ajust analític.
4*e*) Saber calcular les diferents components d'una sèrie, suposant un model additiu, pel mètode de les mitjanes mòbils.
4*f*) Saber interpretar els resultats obtinguts de les diferents components d'una sèrie temporal i relacionar-los amb la gràfica de la sèrie.
4*g*) Fer estimacions dels valors d'una sèrie temporal en dates futures properes als valors analitzats, en l'ajust analític.

Objectiu Exercici	4a	4b	4c	4d	4e	4f	4g
1	x	x	x	x		x	x
2	x	x	x	x		x	x
3	x	x	x	x		x	x
4	x	x	x				
5	x	x	x		x	x	
6	x	x	x		x	x	
7	x	x	x	x	x	x	x

Enunciats

4*a*) Calcular les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual pel mètode de l'ajust analític.
4*b*) Interpretar els resultats obtinguts per a poder explicar les característiques del comportament de la sèrie.
4*c*) Fer estimacions de valors de la sèrie per a períodes immediats.

Exercici 1

Per a analitzar l'evolució de les despeses en un cert departament d'una empresa, es van prendre les dades següents, que expressen en milers d'euros les despeses quadrimestrals dels quatre anys que figuren en la taula:

	2008	2009	2010	2011
1r quadrimestre	26	25	21	20
2n quadrimestre	18	15	12	10
3r quadrimestre	22	20	18	12

a) Suposant un model additiu, calculeu pel mètode de l'ajust analític, les components d'aquesta sèrie i interpreten cadascun dels resultats obtinguts.
b) Estimeu els valors de les despeses que es poden esperar per a l'any 2012.

4*a*) Calcular les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual pel mètode de l'ajust analític.
4*b*) Interpretar els resultats obtinguts per a poder explicar les característiques del comportament de la sèrie.
4*c*) Fer estimacions de valors de la sèrie per a períodes immediats.

Exercici 2

Les dades següents, extretes de l'INI, ens mostren les pernoctacions hoteleres a la Comunitat Valenciana, des de l'any 2006 fins al 2010.

Realitza una anàlisi del fenomen, obtenint les diferents components de la sèrie temporal per l'ajust analític, suposant un model additiu. Interpreteu el significat de cadascuna de les components.

Feu les previsions que podem esperar per els anys 2012 i 2013.

	2006	2007	2008	2009	2010
Gener	1.296.648	1.307.384	1.289.627	1.115.865	1.061.450
Febrer	1.424.453	1.472.806	1.562.471	1.329.013	1.347.885
Març	1.750.282	1.892.925	1.993.175	1.700.733	1.800.323
Abril	2.152.783	2.226.734	1.868.049	1.925.454	1.943.080
Maig	2.131.194	2.181.889	2.114.631	1.952.409	2.049.616
Juny	2.399.782	2.523.555	2.311.444	2.241.147	2.049.616
Juliol	2.884.491	2.983.227	2.877.028	2.871.011	2.945.899
Agost	3.153.407	3.308.833	3.227.055	3.274.561	3.341.961
Setembre	2.540.711	2.580.090	2.524.888	2.381.776	2.424.567
Octubre	2.207.203	2.097.005	2.012.838	1.954.615	2.053.500
Novembre	1.674.433	1.741.296	1.498.967	1.471.579	1.472.610
Desembre	1.437.036	1.420.988	1.251.809	1.201.342	1.174.451

4*a*) Calcular les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual pel mètode de l'ajust analític.
4*b*) Interpretar els resultats obtinguts per a poder explicar les característiques del comportament de la sèrie.
4*c*) Fer estimacions de valors de la sèrie per a períodes immediats.

Exercici 3

Amb les dades següents, extretes de la DGT, que ens mostren les noves llicències de tots els tipus de carnets de conduir a la Comunitat Valenciana, des de l'any 2008 fins al 2010, realitza una anàlisi del fenomen, obtenint les diferents components de la sèrie temporal per l'ajust analític.

Interpreteu el significat de cadascuna de les components.

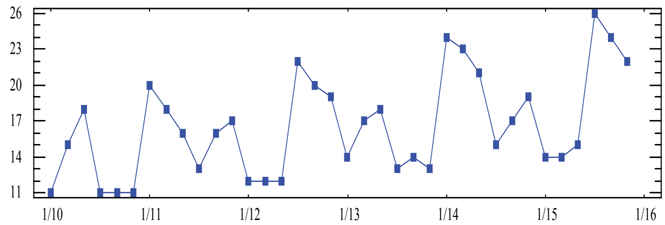
Feu les previsions que podem esperar per els anys 2011 i 2012.

	2008	2009	2010
Gener	12031	8380	7071
Febrer	12208	8993	7685
Març	9497	7973	8444
Abril	12862	7360	6781
Maig	12567	7874	7728
Juny	12723	7881	7585
Juliol	19003	11820	11138
Agost	2147	1346	2205
Setembre	10826	8876	7901
Octubre	11196	8374	7427
Novembre	10628	10137	7665
Desembre	9064	8083	5746

4*a*) Reconèixer en una col·lecció de dades els patrons i notació d'una sèrie temporal.
4*b*) Analitzar una sèrie i a partir de la seua gràfica, poder comprovar que s'adapta al model additiu.
4*c*) Conèixer les diferents components d'una sèrie temporal: tendència, component estacional, component erràtica o residual.

Exercici 4

La gràfica següent és la representació d'una sèrie temporal on es detallen les dades bimensuals de 6 anys. Si haguérem de calcular la tendència i la component estacional de la sèrie pel mètode de les mitjanes mòbils, explica l'elecció del nombre de dades que cal considerar per al càlcul de les mitjanes (p) en cada cas, i justificar la resposta.



4*a*) Calcular les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual pel mètode de les mitjanes mòbils.
4*b*) Interpretar els resultats obtinguts per poder explicar les característiques del comportament de la sèrie.

Exercici 5

En la taula següent presentem el nombre total de viatgers transportats en els serveis de transport públic a la Comunitat Valenciana, detallats per mesos, dels anys 2006 al 2010.

	2006	2007	2008	2009	2010
Gener	12340	12542	12622	11869	10283
Febrer	11850	12156	12202	11365	10981
Març	13721	13729	10457	12024	12207
Abril	10919	11612	12960	10469	10549
Maig	13495	13777	12713	12038	12076
Juny	13029	13094	12619	12314	11783
Juliol	12118	12364	12351	11490	10621
Agost	8803	8814	8846	8027	7698
Setembre	12148	11768	12057	10964	10705
Octubre	13141	13266	13277	11956	11447
Novembre	13307	12655	12474	11718	11557
Desembre	11859	11624	11682	10894	10769

a) Calculeu les components d'aquesta sèrie pel mètode de les mitjanes mòbils.

Interpreteu-ne els resultats.

4*a*) Calcular les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual pel mètode de les mitjanes mòbils.

4*b*) Interpretar els resultats obtinguts per a poder explicar les característiques del comportament de la sèrie.

Exercici 6

Realitzeu la gràfica de la sèrie següent que indica els milers de quilos de fruita comercialitzada per trimestres al llarg dels darrers 4 anys.

	2008	2009	2010	2011
1r trimestre	10	23	12	29
2n trimestre	11	27	11	28
3r trimestre	9	25	10	21
4t trimestre	8	20	8	23

Observeu les taules següents que presenten els càlculs que hem fet per a obtenir la tendència i la component estacional de la sèrie.

Identifiqueu el mètode emprat, afegiu aquelles dades que falten en les taules i comenteu el procediment de càlcul que cal fer hi, justificant-les.

dades		
10		
11		
9		
8		
23		
27		
	16,875	
25		16,9375
	17	
20		17
	17	
12		17,375
11		
	17,875	
10		17,625
	17,375	
8		
29		
28		
21		
23		

dades		
10		
11		
9		
	12,75	
8		
	16,75	
23		18,75
27		
	23,75	
25		22,375
	21	
20		19
	17	
12		15,125
	13,25	
11		11,75
	10,25	
10		
8		16,625
	18,75	
29		20,125
28		
21		
23		

4*a*) i 4*b*) Comparar els dos mètodes que hem treballat: l'ajust analític i les mitjanes mòbils per al càlcul de les components d'una sèrie temporal: tendència, component estacional i component erràtica o residual.
4*c*) Comparar i interpretar els resultats obtinguts per a poder explicar les característiques del comportament de la sèrie.
4*d*) Estimar els valors d'una sèrie temporal en dates futures mitjançant el mètode de l'ajust analític.

Exercici 7

La sèrie cronològica següent mostra el nombre de noves contractacions de certa superfície comercial per quadrimestres en els anys que s'indiquen:

	2007	2008	2009	2010	2011
1r quadrimestre	41	39	35	21	22
2n quadrimestre	37	33	27	15	16
3r quadrimestre	36	30	16	12	13

a) Realitzeu la gràfica de la sèrie i, suposant un model additiu, calculeu pel mètode de l'ajust analític les components d'aquesta sèrie i interpreteu cadascun dels resultats obtinguts.
b) Feu les previsions que podem esperar per els anys 2012 i 2013.
c) Realitzeu l'anàlisi de la sèrie pel mètode de les mitjanes mòbils, calculant també les components de la sèrie.
d) Compareu els resultats de l'anàlisi pels dos mètodes emprats.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de realitzar els exercicis i problemes. És convenient no fer un abús excessiu d'aquestes ajudes, és a dir, abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10-15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2 i en darrer terme la solució.

Ajudes Tipus 1

Exercici 1

Per a resoldre aquest problema seria convenient omplir la taula següent. En la part inferior de les dades realitzarem els càlculs referents a la recta de tendència. Amb els valors de les cel·les dels totals podrem resoldre el sistema corresponent.

Observeu la diferència en aquest apartat pel fet de ser un nombre senar de columnes (a l'exercici 2 hi havia 4 columnes) i per a fer l'escala tal com està indicada en la taula següent (fila i).

En les columnes de la dreta podrem fer els càlculs per a obtenir la component estacional i els índexs estacionals (en%).

	2006	2007	2008	2009	2010		$\bar{y}_k$	$\bar{y}_k'$	e_k	I_k
						TOTALS				
$\bar{y}_i$								M =		
i	-2	-1	0	1	2					
$\bar{y}_i \cdot i$										
i^2										

Recordeu interpretar els resultats que us permetran confirmar la coherència dels resultats amb les dades originals.

Com que es tracta d'una sèrie mensual amb moltes dades caldria ajudar-se amb un full de càlcul.

Per a fer les previsions, caldrà calcular prèviament la tendència de cada any i la component de cada període, per a obtenir els resultats que volem, sumarem en cada cas les dades corresponents.

Exercici 2

Per a fer aquest exercici seguirem el mateix procediment que en l'exercici 2. També es tracta d'una sèrie mensual per la qual cosa seria convenient utilitzar un full de càlcul. Per tractar-se també d'un nombre senar de columnes i utilitzant la escala de la fila *i*, posant el 0 a la columna central, se simplifica molt la resolució del sistema que cal plantejar per a la tendència.

Tal com s'ha comentat a l'exercici 2, és convenient treballar les dades i emplenar una taula semblant a la de l'exercici anterior.

Exercici 3

En aquest exercici no es demana fer cap càlcul. Tan sols és una reflexió per aprendre a triar el nombre de dades (p) que cal agafar per a fer les mitjanes mòbils, tan per al càlcul de la tendència com per al càlcul de la component estacional.

Exercici 4

Es tracta d'una sèrie amb moltes dades, per la qual cosa caldria ajudar-se d'un full de càlcul ja que devem abordar l'anàlisi pel mètode de les mitjanes mòbils.

Cal fer, en primer lloc, la gràfica de les dades i buscar-hi la periodicitat. En aquest cas té un comportament que es repeteix any rere any, caldrà doncs calcular mitjanes mòbils de 12 dades ($p = 12$) per al càlcul de la tendència. Per tractar-se d'un nombre parell després haurem de fer un posterior «centrat».

Per a calcular la component estacional podrem aprofitar les mateixes mitjanes de l'apartat anterior, per tractar-se de $p = 12$.

Per a treballar més clarament seria convenient preparar una taula semblant a aquesta (presentem tan sols els primers càlculs i espais).

Les gràfiques poden servir-nos també per a la interpretació dels resultats i la validació de la seua coherència amb les dades originals.

Gener 2006	12.340		
Febrer 2006	11.850		
Març 2006	13.721		
Abril 2006	10.919		
Maig 2006	13.495		
Juny 2006	13.029		
		12227,5	
Juliol 2006	12.118		12235,9167
		12244,3333	
Agost 2006	8.803		12257,0833
		12269,8333	

Exercici 5

En aquest exercici ja ens donen les taules preparades per al mètode de les mitjanes mòbils i amb molts càlculs ja realitzats. És un exercici de consolidació de la tècnica i també es treballa el mateix objectiu del problema 4, que és aprendre a esbrinar el nombre de dades que cal agafar per a calcular la tendència i la component estacional. Comença pensant quin és aquest nombre *p* en cada cas i completa les taules on falten valors al principi, al mig i al final. Per tractar-se de nombres parells cal fer després un «centrat» en cada cas.

Tot seguit s'ha de calcular la component residual.

Exercici 6

En realitat aquest exercici es «doble» ja que ens demanen l'anàlisi pels dos mètodes. A més, ens servirà per acabar el tema i consolidar el nostre treball repassant les dues tècniques d'anàlisi que hem vist al llarg del tema.

Caldrà fer, en primer lloc, una anàlisi pel mètode de l'ajust analític. Podeu seguir la pauta dels exercicis 2 i 3 en aquest apartat.

Per a fer l'anàlisi pel mètode de les mitjanes mòbils podeu seguir la pauta dels exercicis 5 i 6 però amb menys dades, la qual cosa simplificarà els càlculs.

Ens demanen que comparem ambdós mètodes per a reflexionar sobre les semblances i diferències entre ells.

Ajudes Tipus 2

Exercici 1

Per a realitzar l'anàlisi d'una sèrie temporal pel mètode de l'ajust analític és convenient disposar les dades en la taula següent, com ja s'indicava en la de l'ajuda tipus 1.

	06	07	08	09	10		$\bar{y}_k$	$\bar{y}'_k$	e_k	I_k
						TOTALS				
$\bar{y}_i$						10200467		M = 2.059.542		
i	-2	-1	0	1	2	0				
$\bar{y}_i \times i$						-424346				
i^2	4	1	0	1	4	10				

Per a calcular la tendència, ens fixarem en les cel·les de la part inferior. Hem omplit els valors de la columna dels «totals» per a poder comprovar els vostres càlculs i a més resoldre el sistema següent:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases} \Rightarrow \begin{cases} 10200467 = 5a + 0b \\ -424346 = 0a + 10b \end{cases}$$

Els resultats d'aquest sistema ens permet obtenir la recta de tendència:

$$T_i = a + b \;\; = 2040093{,}4 - \; 42434{,}6i$$

Per a calcular la component estacional, ens fixarem en les cel·les de la part de la dreta. Recordem que cal determinar *b/m*= –42434,6/12 = –3536,2 per a poder obtenir les mitjanes corregides. Indiquem el valor de *M* a la taula.

La component estacional la obtenim en valors absoluts i com a índex, on el valor de *M* és el valor de referència.

Per a obtenir la component residual $r_{ik} = y_k - T_i - e_k$, per la qual cosa cal tenir prèviament calculats els valors de la tendència (per a cada any) i el valor de la component estacional (per a cada període).

Per a fer les previsions s'ha de calcular la tendència que correspondrà als anys 2012 i 2013, i construir-hi la sèrie per als períodes que ens demanen sumant aquestes components: $y_{ik} = T_i + e_k$.

Exercici 2

Aquests exercici és molt semblant a l'anterior i la taula i els càlculs seguiran el mateix plantejament.

La taula a emplenar amb els «totals» serà la següent:

	08	09	10		$\overline{y}_k$	$\overline{y}_k^{'}$	e_k	I_k
				TOTALS		M = 9.772		
$\overline{y}_i$				26602				
i	-1	0	1	0				
$\overline{y}_i \times i$				-3948				
i^2	1	0	1	2				

Alguns resultats parcials que podeu comprovar:

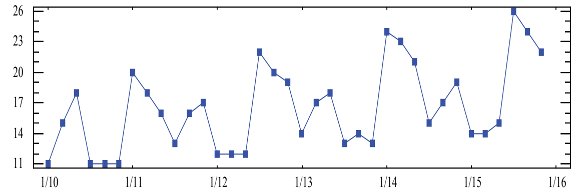
Recta de tendència: $T_i = a + bi = 8867{,}36 - 1974i$
b/m= $-164{,}5$

$e_1 = -611$	$e_2 = 21$	$e_3 = -805$
$e_4 = -278$	$e_5 = 276$	$e_6 = 447$
$e_7 = 5.202$	$e_8 = -6.721$	$e_9 = 745$
$e_{10} = 707$	$e_{11} = 1.350$	$e_{12} = -332$

Amb aquests valors cal trobar la component erràtica o residual i les previsions que ens demanen.

Exercici 3

Per a resoldre aquest exercici és necessari observar la gràfica de la sèrie i el nombre de dades que hi ha en cada «període» de la gràfica.



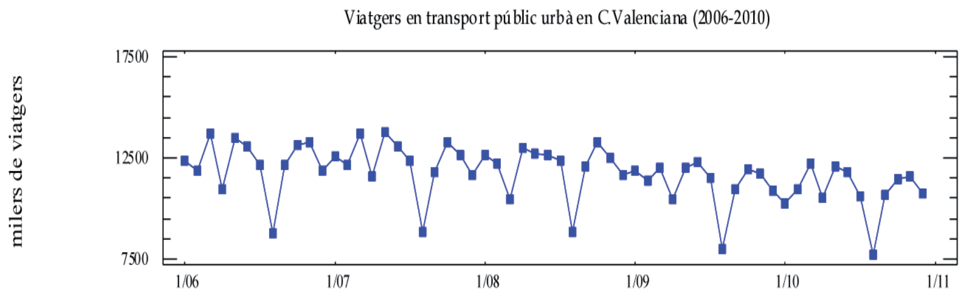
En aquest cas, per a calcular la tendència devem considerar que en aquesta gràfica es «repeteix» el patró cada 9 observacions, i que per tractar-se de dades bimensuals tenim 6 dades per any. Quin serà, doncs, el nombre de dades que cal agafar en cada mitjana per a fer un «suavitzat» pel mètode de les mitjanes mòbils?

Per a calcular la component estacional, quin serà el nombre de dades a agafar? Quantes dades tenim per any?

Exercici 4

Es tracta d'una sèrie amb moltes dades per la qual cosa cal advertir que seria recomanable fer els càlculs amb l'ajut d'un full de càlcul.

Com que utilitzarem el mètode de les mitjanes mòbils, en primer lloc caldrá decidir el nombre de dades que s'ha de considerar en cada mitjana, per la qual cosa cal veure la gràfica de la sèrie:



Podem veure que té una periodicitat anual i com que tenim dades mensuals, doncs s'han de calcular les mitjanes amb 12 dades. Per tractar-se d'un nombre parell de dades, després s'haurà de fer un «centrat».

En la taula següent, indiquem els primers resultats que podreu comprovar:

Gener 2006	12340		
Febrer 2006	11850		
Març 2006	13721		
Abril 2006	10919		
Maig 2006	13495		
Juny 2006	13029		
		12227,5	
Juliol 2006	12118		12235,9167
		12244,3333	
Agost 2006	8803		12257,0833
		12269,8333	
Setembre 2006	12148		12270,1667
		12270,5	
Octubre 2006	13141		12299,375
		12328,25	
Novembre 2006	13307		12340
		12351,75	
Desembre 2006	11859		12354,4583
		12357,1667	
Gener 2007	12542		
		....	
Febrer 2007	12.156		....
		....	

La columna de la dreta recull els valors de la tendència.

Per a determinar la component estacional, calculem les mitjanes de les dades originals, per períodes i les indiquem amb $\bar{y}_k$:

	2006	2007	2008	2009	2010	$\bar{y}_k$
Gener	12340	12542	12622	11869	10283	
Febrer	11850	12156	12202	11365	10981	
Març	13721	13729	10457	12024	12207	
Abril	10919	11612	12960	10469	10549	
Maig	13495	13777	12713	12038	12076	
Juny	13029	13094	12619	12314	11783	
Juliol	12118	12364	12351	11490	10621	
Agost	8803	8814	8846	8027	7698	
Setembre	12148	11768	12057	10964	10705	
Octubre	13141	13266	13277	11956	11447	
Novembre	13307	12655	12474	11718	11557	
Desembre	11859	11624	11682	10894	10769	

Per altra banda, en aquest cas, podem aprofitar els resultats de les mitjanes de la tendència (per estar calculades amb 12 dades, observacions que tenim de cada any), encara que cal redistribuir-los en anys i períodes, en forma de taula i calcular les seues mitjanes per períodes (files). Les indicarem amb E_k.

A la taula següent teniu alguns valors per comprovar, encara que hi ha d'altres que s'hauran de calcular.

	2006	2007	2008	2009	2010	E_k
Gener					11041,125	11776,9896
Febrer					10991,2083	11749,8854
Març					10966,7083	11723,3437
Abril					10934,7083	11690,6667
Maig					10906,7917	11654,7917
Juny					10894,875	

Juliol	12235,9167					
Agost	12257,0833					
Setembre	12270,1667					
Octubre	12299,375					
Novembre	12340					
Desembre	12354,4583					

Per a calcular la component estacional caldrà restar les columnes finals d'aquestes dues taules, $e_k = \bar{y}_k - E_k$, cosa que ens permetrà explicar quins són els mesos de major i menor utilització del transport públic a la Comunitat Valenciana.

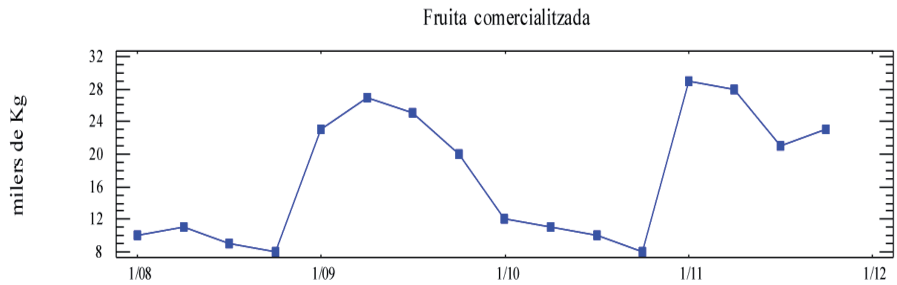
Per a determinar la component erràtica o residual cal restar a cada valor inicial el valor de la tendència corresponent a cada cel·la en els casos en què aquesta existeix (no als primers ni als darrers valors de la sèrie), i també li restem el valor de la component estacional que correspon a cada període.

En la taula següent es presenten els resultats d'aquests càlculs que podeu comprovar. Sabrieu explicar què ens diuen els valors ressenyats en roig i en blau?

	2006	2007	2008	2009	2010
Gener		20,3729	462,9979	20,1646	-912,335
Febrer		-183,0396	235,5021	-220,5396	28,8771
Març		661,993725	-2266,21458	-225,214575	536,035425
Abril		-351,25835	1329,90835	-586,50835	3,15835
Maig		281,82497	-476,38333	-484,84163	4,19997
Juny		-141,799975	-342,841675	77,908325	-54,466675
Juliol	20,168725	215,335425	498,793725	433,502125	
Agost	-77,82652	-101,74322	298,21508	290,75678	
Setembre	106,939145	-160,477455	331,647545	88,980845	
Octubre	-59,283964	287,674336	460,132736	-59,992264	
Novembre	302,245684	-99,004316	25,203984	-66,754316	
Desembre	-224,748979	-130,415679	209,501021	65,251021	

Exercici 5

En l'enunciat d'aquest exercici ens presenten dues taules que, evidentment, es corresponen al mètode de les mitjanes mòbils.



Mirem la gràfica de la sèrie (el patró de periodicitat es pot dir que es repeteix cada dos anys, 8 observacions) i com que es tracta de dades trimestrals (4 dades per any), hem d'agafar 8 dades per a calcular les mitjanes de la tendència, ja que MCM(8,4) = 8.

En la taula de l'esquerra falten valors per omplir al començament, al mig i a l'acabament. També falten per calcular cel·les de la columna de la dreta que correspon al «centrat» (mitjana aritmètica de cada dos valors de la columna anterior).

Mentrestant, en la taula de la dreta que ens proposen a l'exercici, les mitjanes que s'han calculat són de 4 observacions, ja que per tractar-se de dades trimestrals, en tenim 4 per any. Podem comprovar que falten els resultats d'algunes cel·les en la columna de les mitjanes com la del «centrat» de la columna de la dreta. Aquestes mitjanes ens permetran calcular la component estacional.

Caldrà repetir el procés del problema anterior:

a) Calcular les mitjanes de les dades originals per períodes $\bar{y}_k$.
b) Organitzarem en forma de taula de doble entrada (anys i trimestres) els resultats de la darrera columna de la taula de la dreta.
c) Calcularem les mitjanes per a cada període de la taula anterior, E_k.
d) Els valors de la component estacional s'obtenen restant: $e_k = \bar{y}_k - E_k$.
e) Interpretarem els resultats.

Per a calcular la component residual o erràtica: $r_k = y_k - T_k - e_k$. Veurem que tan sols podem determinar-la per als dos anys centrals per no disposar de més dades de tendència.

Exercici 6

Aquest exercici proposa fer l'anàlisi de la sèrie pels dos mètodes per a convidar-nos a confrontar les dues tècniques.

a) Pel mètode de l'ajust analític:

Caldrà emplenar la següent taula. Per al càlcul de la tendència necessitarem les cel-les de la part inferior i utilitzarem la columna de «totals» per a plantejar el sistema que ens permet obtenir els coeficients de la recta de tendència:

	2007	2008	2009	2010	2011		$\overline{y}_k$	$\overline{y}_k^{'}$	e_k	I_k
1r quad.	41	39	35	21	22					112,06%
2n quad.	37	33	27	15	16					97,87%
3r quad.	36	30	16	12	13	TOTALS ↓				90,07%
$\overline{y}_i$	38	34	26	16	17	131		M = 28,2		
i						0				
$\overline{y}_i \times i$						-60				
i^2						10				

En la taula presentem alguns resultats que podrem comprovar i, a continuació, els resultats que cal obtenir amb els valors de la taula:

Recta de tendència: $T_i = 26,-6i$

$$b/m = -6/3 = -2$$

$$M = \frac{\sum_k \overline{y}_k^{'}}{m} = 28,2$$

Component estacional:

$$e_1 = \overline{y}_1^{'} - M = 31,6 - 28,2 = 3,4$$
$$e_2 = \overline{y}_2^{'} - M = 27,6 - 28,2 = -0,6$$
$$e_3 = \overline{y}_3^{'} - M = 25,4 - 28,2 = -2,8$$

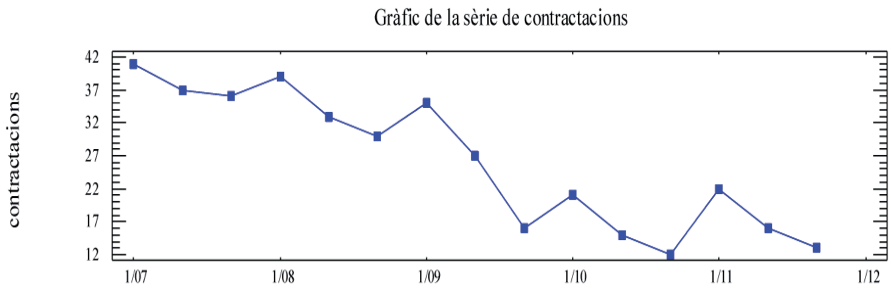
Índexs estacionals:

$$I_1 = \frac{\overline{y}_1^{'}}{M} \cdot 100 = \frac{31,6}{28,2} \cdot 100 = 112,05\%$$
$$I_2 = \frac{\overline{y}_2^{'}}{M} \cdot 100 = \frac{27,6}{28,2} \cdot 100 = 97,87\%$$
$$I_3 = \frac{\overline{y}_3^{'}}{M} \cdot 100 = \frac{25,4}{28,2} \cdot 100 = 90,07\%$$

Per a calcular la component residual: $r_{ik} = y_{ik} - T_i - e_k$

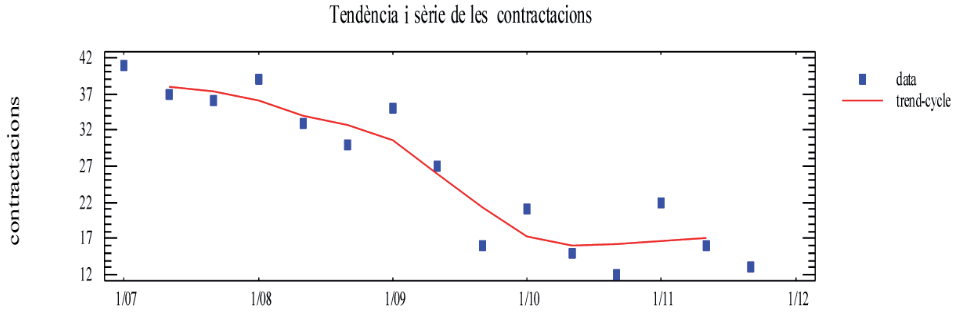
b) Pel mètode de les mitjanes mòbils:

Cal veure la gràfica de la sèrie i com que el seu comportament és anual i tenim dades quadrimestrals, agafarem 3 dades per a calcular cada mitjana i obtenir la columna de les mitjanes mòbils. Com que es tracta d'un nombre senar de dades no cal fer un posterior «centrat».



En la taula següent figuren alguns resultats dels valors per comprovar. Falten altres cel·les per omplir. La columna, ja completada, són els valors de la tendència. Pots ajudar-te d'aquest gràfic per a interpretar-la:

dades	T_k
41	
37	38,00
36	37,33
39	
33	
30	32,67
35	3067
27	
16	
21	17,33
15	16,00
12	
22	
16	17,00
13	



Amb aquests resultats, que hem obtingut de les mitjanes, els col·loquem de nou en forma de taula de doble entrada (com a l'enunciat) fent correspondre a cada període de la taula la seua mitjana i, de nou, repetirem tot el procediment com hem indicat a l'exercici 6.

La comparació dels dos mètodes podeu trobar-la donant resposta a aquestes qüestions entre d'altres possibles que hi podeu afegir:

a) Els resultats son iguals? I la seua interpretació?
b) Són igualment fiables o la qualitat de les interpretacions depén de les característiques de les dades (periodicitat, regularitat, etc.)?
c) Podeu fer previsions en ambdós mètodes?

L'exercici que heu d'haver trobat és el núm. 6.

Les portades es diferencien únicament per les notícies que hi apareixen, i no pas per la posició de cada notícia en la primera plana. Per tant, cal comptar combinacions de 6 notícies agrupades de 4 en 4.

Solucions

Exercici 1

Per a analitzar l'evolució de les despeses en un cert departament d'una empresa, es van prendre les següents dades, que expressen en milers d'euros les despeses quadrimestrals dels quatre anys que figuren en la taula:

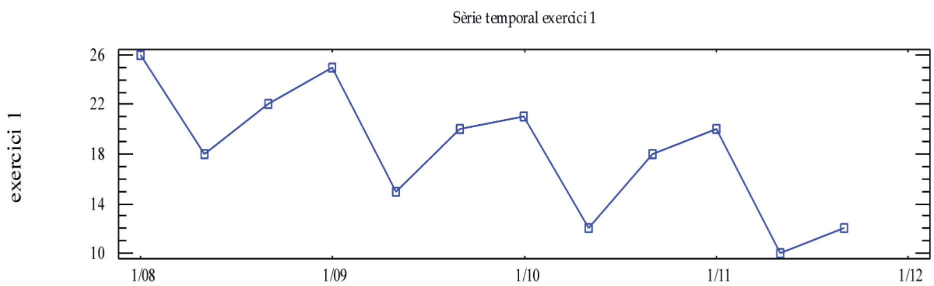
	2008	2009	2010	2011
1r quadrimestre	26	25	21	20
2n quadrimestre	18	15	12	10
3r quadrimestre	22	20	18	12

Solució

a) Suposant un model additiu, calculeu pel mètode de l'ajust analític, les components d'aquesta sèrie i interpreteu cadascun dels resultats obtinguts.

Identificació del model i gràfica

Sempre cal començar amb una gràfica de les dades per veure el patró de comportament de la sèrie, i així confirmar que podem aplicar-li un ajust analític de tipus additiu:



Tan sols mirar aquesta gràfica cal esperar una tendència decreixent i també es veu una component estacional prou marcada per la «periodicitat» que podem veure en la gràfica (eix X) que coincideix amb l'interval anual que correspon cada tres dades.

Càlcul de la tendència

Per a calcular la component de la tendència emplenarem les caselles de la part inferior de la taula següent, per a ajustar una recta de regressió a les mitjanes quadrimestrals anuals que figuren en la fila $\overline{y}_i = \frac{\sum_k y_{ik}}{m}$.

Cal notar que $\overline{y}_i$ és la mitjana dels valors de cada columna. Recordem tan sols que cada dada d'una sèrie temporal la representem per dos subíndexs y_{ik} on i fa referència a l'any i k fa referència al període, m indica el nombre de files de la taula o nombre de períodes dels quals tenim dades per cada any. En el nostre cas $m = 3$ perquè un any té 3 quadrimestres.

Veurem que en la fila següent podem trobar una escala i que fa referència als anys per simplificar els càlculs; és recomanable posar el valor 0 en alguna de les columnes centrals de la taula perquè simplificarà molt els càlculs posteriors.

A continuació emplenarem les files tercera i quarta que fan referència a $\overline{y}_i \cdot i$ i i^2, considerant els valors de les files anteriors.

Després sumarem les files i obtindrem els valors que podem trobar en la columna «totals».

	2008	2009	2010	2011	
1r quadrimestre	26	25	21	20	
2n quadrimestre	18	15	12	10	
3r quadrimestre	22	20	18	12	TOTALS
$\overline{y}_i$	22	20	17	14	73
i	-1	0	1	2	2
$\overline{y}_i \times i$	-22	0	17	28	23
i^2	1	0	1	4	6

Utilitzarem aquests valors «totals» per a resoldre el sistema que es planteja a continuació:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases}$$

on N és el nombre d'anys dels quals tenim dades a la taula (en aquest cas, 4 anys que corresponen a les 4 columnes) i els coeficients a i b són els coeficients de la recta de regressió que denominem recta de tendència, la fórmula de la qual és $T_i = a + b$, on i fa referència a l'any que se indica en l'escala de la taula superior.

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases} \Rightarrow \begin{cases} 73 = 4a + 2b \\ 23 = 2a + 6b \end{cases}$$

Resoldrem el sistema pel mètode que considerem més adient (seria molt fàcil per reducció) i trobem els valors a = 19,6 i b = –2,7, amb els quals podem concloure que T_i = 19,6 – 2,7i.

Per a interpretar aquesta component de la sèrie, T_i, que s'anomena tendència i que ens explica el comportament del fenomen a «llarg termini», ens fixem en el valor de la pendent b = 2,7, que per ser negativa ens permet explicar que les despeses del departament van decreixent any rere any, i podem a més detallar que el valor de la mitjana de despeses quadrimestrals del departament, $\overline{y}_i$, cada any ha disminuït en 2,7 milers d'euros.

Càlcul de la component estacional

La segona component de la sèrie que cal determinar és la component estacional, que denotarem per e_k, que ens permetrà analitzar el comportament del fenomen per períodes dintre de l'any (en el nostre exercici, per quadrimestres), valorant en quals d'ells els valors estan per damunt o per sota d'un valor global que denotarem per M i que anomenarem mitjana global corregida.

Els valors de e_k estaran expressats en valors absoluts i en les mateixes unitats que les dades de la taula original (en aquest exercici en milers d'euros).

Per a abordar els càlculs de e_k, ampliarem la taula en diferents columnes cap a la dreta de l'original, ja que ens interessa fer un treball per períodes, en aquest cas, per quadrimestres.

En la columna primera calcularem la mitjana dels valors de cada quadrimestre i aquesta mitjana la denotarem per $\overline{y}_k = \frac{\sum_i y_{ij}}{N}$.

	2008	2009	2010	2011		$\overline{y}_k$	$\overline{y}_k^{'}$	e_k	I_k
1r quadrimestre	26	25	21	20		23	23	3,85	120,10%
2n quadrimestre	18	15	12	10		13,75	14,65	-4,5	76,50%
3r quadrimestre	22	20	18	12	TOTALS ↓	18	19,8	0,65	103,39%
							M = 19,15		

En la següent columna calcularem les mitjanes corregides $\overline{y}_k^{'}$ (on eliminem en cada dada el valor proporcional de la tendència que podem assignar a cada període) suposant que aquest decreixement $b = -2,7$ ha estat constant al llarg dels períodes de l'any.

En aquest exercici, $b/m = -2,7/3 = -0,9$. En ser una tendència negativa, els valors de les mitjanes corregides $\overline{y}_k^{'}$ seran majors que les mitjanes originals $\overline{y}_k$, ja que aquestes es calculen així:

$$\overline{y}_k^{'} = \overline{y}_k - \frac{b}{m}(k-1)$$

Així:

$$\overline{y}_1^{'} = \overline{y}_1 - \frac{b}{m}(1-1) = 23-(-0,9)(1-1) = 23$$

$$\overline{y}_2^{'} = \overline{y}_2 - \frac{b}{m}(2-1) = 13,75-(-0,9)(2-1) = 14,65$$

$$\overline{y}_3^{'} = \overline{y}_3 - \frac{b}{m}(3-1) = 18-(-0,9)(3-1) = 19,8$$

A continuació, en una cel·la inferior, calculem M, la mitjana de les mitjanes corregides, ja que:

$$M = \frac{\sum_k \overline{y}_k^{'}}{m} = \frac{23+14,65+19.8}{3} = 19,15$$

i representa el valor de referència per analitzar els valors dels períodes, mitjançant $e_k = \overline{y}_k^{'} - M$, que calcularem en la columna següent.

$$e_1 = \overline{y}_1^{'} - M = 23-19,15 = 3,85$$

$$e_2 = \overline{y}_2^{'} - M = 14,65 - 19,15 = -4,5$$
$$e_3 = \overline{y}_3^{'} - M = 19,8 - 19,15 = 0,65$$

La component estacional e_k ens explicita el comportament per períodes en valors absoluts, indicant-nos signe i quantitat en les mateixes unitats que les dades originals (milers d'euros). Així, direm que el primer quadrimestre les despeses del departament de la nostra empresa tenen un valor per damunt dels 3850 euros, mentre que les despeses del segon quadrimestre són 4500 euros per sota de la mitjana i el tercer quadrimestre son superiors tan sols en 650 euros a dita mitjana. El valor de la mitjana de referència seria M = 19150 euros que podria representar una mitjana de despeses quadrimestrals global.

Una altra interpretació d'aquestes dades es pot donar amb els índexs estacionals $I_k = \frac{\overline{y}_k}{M} \cdot 100$ que calculem en la següent columna, on s'indica en forma de percentatge (valor relatiu, on M representa el 100%) la mateixa informació que la component estacional, però que en tenir caràcter percentual es més fàcil de presentar sense particularitzar i donar el valor de M.

$$I_1 = \frac{\overline{y}_1^{'}}{M} \cdot 100 = \frac{23}{19,15} \cdot 100 = 120,10\%$$
$$I_2 = \frac{\overline{y}_2^{'}}{M} \cdot 100 = \frac{14,65}{19,15} \cdot 100 = 76,50\%$$
$$I_3 = \frac{\overline{y}_3^{'}}{M} \cdot 100 = \frac{19,8}{19,15} \cdot 100 = 103,39\%$$

Aquesta component estacional expressada en termes de percentatge, els índexs estacionals, ens permet analitzar i interpretar el comportament de les dades de la sèrie, és a dir, els valors de les despeses del departament, per quadrimestres.

Es pot afirmar que les despeses eren majors al primer quadrimestre, amb valors un 20,10% superiors a la mitjana anual, mentre que els valors del tercer quadrimestre tan sols superen dit valor en un 3,39%. Cal destacar que les despeses disminueixen al segon quadrimestre amb valors que estan al voltant del 76,50% del valor de dita mitjana anual global, que podríem considerar al valor de M = 19,15 milers d'euros.

Càlcul de la component residual o erràtica

En tercer lloc, cal determinar la component erràtica o residual, que denotem per r_{ik}, que ens permet destacar el comportament d'alguna dada y_{ik}, el valor de la qual no es puga explicar per les anteriors components, la qual cosa permetrà inferir que per qualsevol causa a identificar (motius extraordinaris) aquest valor no està dintre del patró de comportament que hem trobat i amb el qual hem interpretat les dades originals per explicar el fenomen. Els valors d'aquesta component erràtica cal que siguen menuts en valor, variats en signe i sense cap regularitat ni patró. Ens permeten veure que les dades reals no s'ajusten completament al patró que hem trobat amb la tendència i la component estacional. Per això si algun valor es molt alt o baix deixarà identificar alguna dada que corresponga al període d'un any, el valor de la qual s'allunya molt dels valors que caldria esperar.

La calcularem així $r_{ik} = y_{ik} - T_i - e_k$ en cada cel·la de la taula. Per a tal fi, primerament cal determinar la tendència per a cada any de la taula $T_i = 19,6 - 2,7i$.

$$T_{2008} = T_{-1} = 19,6 - 2,7(-1) = 22,3$$
$$T_{2009} = T_0 = 19,6 - 2,7 \cdot 0 = 19,6$$
$$T_{2010} = T_1 = 19,6 - 2,7 \cdot 1 = 16,9$$
$$T_{2011} = T_2 = 19,6 - 2,7 \cdot 2 = 14,2$$

Els valors de la component estacional estan en la columna de la taula e_k, Així:

$$e_1 = 3,85 \qquad e_2 = -4,5 \qquad e_3 = 0,65$$

Així, els càlculs per a cada cel·la estan en la següent taula:

	2008	2009	2010	2011
1r quadrimestre	26 – 22,3 – 3,85 = –0,15	25 – 19,6 – 3,85 = 1,55	21 – 16,9 – 3,85 = 0,25	20 – 14,2 – 3,85 = 1,95
2n quadrimestre	18 – 22,3 + 4,5= 0,2	15 – 19,6 + 4,5 = –0,1	12 – 16,9 + 4,5 = –0,4	10 – 14,2 + 4,5 = 0,3
3r quadrimestre	22 – 22,3 – 0,65 = –0,95	20 – 19,6 –0,65 = –0,25	18 – 16,9 – 0,65= 0,45	12 – 14,2 – 0,65 = –2,85

D'aquests resultats, podem observar que els valors que pot ser ens criden més l'atenció serien els corresponents al primer quadrimestre de 2011, que es superior al que caldria esperar amb $r_{2011,1} = 1,95$, i el tercer quadrimestre del mateix any amb un valor encara més inferior amb $r_{2011,3} = -2,85$. Caldria estudiar si alguna circumstància extraordinària justifica aquests valors o pot ser indica que el comportament del fenomen està canviant substancialment. Caldria estar atents a les properes dades de l'any següent.

b) Estimeu els valors de les despeses que es poden esperar per a l'any 2012.

Per a fer les estimacions que ens demanen en aquest apartat, recordem que no podem preveure la component residual, per la qual cosa calcularem:

$$y_{ik} = T_i + e_k$$

Considerant l'escala que hem adoptat en la taula per als anys, amb el propòsit de simplificar els càlculs de la tendència, podem establir que si $2009 \Rightarrow i = 0$, això comporta $2012 \Rightarrow i = 3$, la qual cosa ens permetrà obtenir el valor de la tendència per a aquest any: $T_{2012} = T_3 = 19,6 - 2,7 \cdot 3 = 11,5$ i com que coneixem la component estacional $e_1 = 3,85$, $e_2 = -4,5$ i $e_3 = 0,65$, podrem estimar els valors de les despeses de l'any 2012, per quadrimestres.

$y_{2012,1qua} = 11,5 + 3,85 = 15,35$ milers d'euros.

$y_{2012,2qua} = 11,5 - 4,5 = 7$ milers d'euros.

$y_{2012,3qua} = 11,5 + 0,65 = 12,15$ milers d'euros.

Exercici 2

Amb les dades següents, extretes de l'INI, es mostren les pernoctacions hoteleres a la Comunitat Valenciana, des de l'any 2006 fins al 2010.

Realitzeu una anàlisi del fenomen, i obteniu les diferents components de la sèrie temporal per l'ajust analític, suposant un model additiu. Interpreteu el significat de cadascuna de les components.

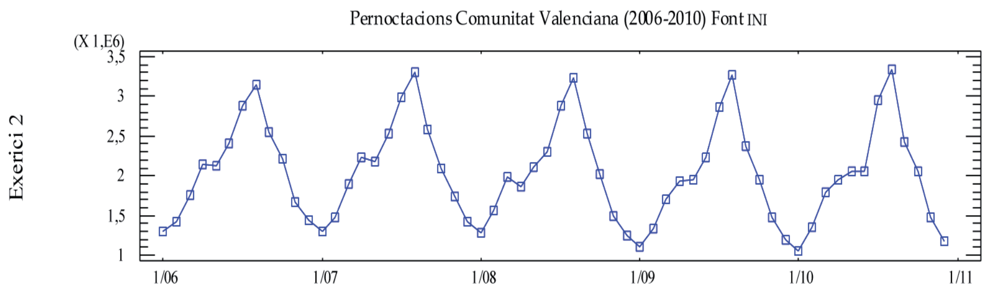
Feu les previsions que podem esperar per els anys 2012 i 2013.

	2006	2007	2008	2009	2010
Gener	1296648	1307384	1289627	1115865	1061450
Febrer	1424453	1472806	1562471	1329013	1347885
Març	1750282	1892925	1993175	1700733	1800323
Abril	2152783	2226734	1868049	1925454	1943080
Maig	2131194	2181889	2114631	1952409	2049616
Juny	2399782	2523555	2311444	2241147	2049616
Juliol	2884491	2983227	2877028	2871011	2945899
Agost	3153407	3308833	3227055	3274561	3341961
Setembre	2540711	2580090	2524888	2381776	2424567
Octubre	2207203	2097005	2012838	1954615	2053500
Novembre	1674433	1741296	1498967	1471579	1472610
Desembre	1437036	1420988	1251809	1201342	1174451

Solució

Anàlisi del model i gràfica

Per començar, cal observar la gràfica de les dades i comprovar que poden ser analitzades pel mètode de l'ajust analític, suposant un model additiu. És fàcil observar un patró que es repeteix amb prou regularitat any rere any, amb un marcat comportament estacional que es repeteix al llarg de la col·lecció de dades que presentem. Per contra, no s'hi observa una tendència de marcada pendent, ja que les dades es mantenen prou constants a llarg termini.



Càlcul de la tendència

Considerarem les dades de la sèrie següent que fa referència a les pernoctacions mensuals hoteleres en la Comunitat Valenciana en els darrers cinc anys:

	2006	2007	2008	2009	2010	
Gener	1296648	1307384	1289627	1115865	1061450	
Febrer	1424453	1472806	1562471	1329013	1347885	
Març	1750282	1892925	1993175	1700733	1800323	
Abril	2152783	2226734	1868049	1925454	1943080	
Maig	2131194	2181889	2114631	1952409	2049616	
Juny	2399782	2523555	2311444	2241147	2049616	
Juliol	2884491	2983227	2877028	2871011	2945899	
Agost	3153407	3308833	3227055	3274561	3341961	
Setembre	2540711	2580090	2524888	2381776	2424567	
Octubre	2207203	2097005	2012838	1954615	2053500	
Novembre	1674433	1741296	1498967	1471579	1472610	

Desembre	1437036	1420988	1251809	1201342	1174451	TOTALS
$\overline{y}_i$	2087702	2144728	2044332	1951625	1972080	10200467
i	-2	-1	0	1	2	0
$\overline{y}_i \cdot i$	-4175404	-2144728	0	1951625	3944160	-424346
i^2	4	1	0	1	4	10

Podem veure que a la part inferior de la taula anterior hem calculat les mitjanes mensuals de cada any $\overline{y}_i$, els valors de l'escala i, i les dues files inferiors $\overline{y}_i \cdot i$, i^2, que ens permetrà calcular en la columna *totals* la suma dels valors de cada fila. En tots els càlculs hem arrodonit a nombres enters per tractar-se de nombre de pernoctacions.

Amb aquestes dades, plantejarem el sistema següent:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases} \Rightarrow \begin{cases} 10200467 = 5a + 0b \\ -424346 = 0a + 10b \end{cases}$$

on els coeficients a i b són els coeficients de la recta de regressió que denominem recta de tendència, la formula de la qual és $T_i = a + bi$, on i fa referència a l'any que s'indica a l'escala de la taula superior.

Podem veure en aquest exemple que l'estratègia de donar-li el valor $i = 0$ a l'any corresponent a la columna central, quan N (nombre d'anys) és imparell simplifica molt la resolució del sistema. Així:

$$a = \frac{10200467}{5} = 2040093{,}4 \qquad b = \frac{-424346}{10} = -42434{,}6$$

són els coeficients de $T_i = a + bi = 2040093{,}4 - 42434{,}6i$ que és l'equació de la recta de tendència i ens permet interpretar que a llarg termini, a partit d'aquestes dades, la mitjana mensual del nombre de pernoctacions disminueix en 42435 pernoctacions per any. Per tractar-se d'una quantitat menuda en relació a les dades de la sèrie és la raó per la qual a la gràfica no hi observàvem clarament el decreixement.

Càlcul de la component estacional

Per a calcular la component estacional, treballarem les columnes que es poden veure a la dreta de la taula següent.

Recordem que en la primera columna, $\overline{y}_k = \frac{\sum_i y_{ij}}{N}$ són les mitjanes de les pernoctacions de cada mes, calculades amb les dades originals, per files.

La columna següent correspon a les mitjanes corregides, la fórmula de les quals és $\overline{y}_k^{'} = \overline{y}_k - \frac{b}{m}(k-1)$ on *b/m* = $-$ 42434,6 /12 = –3536,2; recordem que *m* = 12 perquè, en tractar-se de dades mensuals, en tenim 12 per any.

M és la mitjana global corregida i és la mitjana de les mitjanes corregides:

$$M = \frac{\sum_k \overline{y}_k^{'}}{m}$$

que podem interpretar com el valor mitjà de les pernoctacions mensuals i que representarà el 100% o valor de referència.

Després, les dues columnes següents mostren els valors de la component estacional $e_k = \overline{y}_k^{'} - M$ i els índexs estacionals $I_k = \frac{\overline{y}_k^{'}}{M} \cdot 100$ on podem veure en quins mesos el nombre de pernoctacions està per damunt o per davall del valor de M. e_k ens presenta aquesta desviació en quantitats absolutes (nombre de pernoctacions) mentre que I_k ho indica de manera percentual.

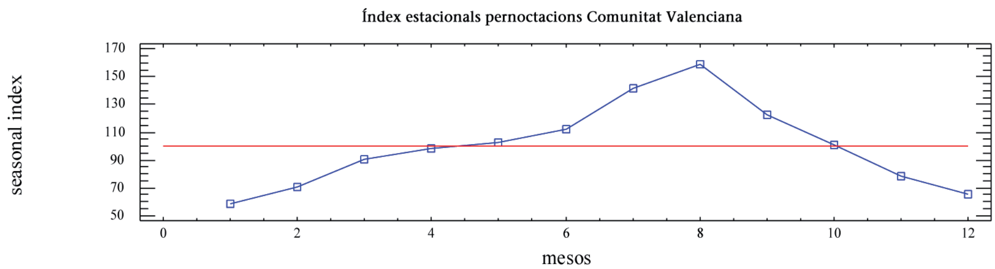
Podem comentar que el mes amb més pernoctacions és agost, amb un valor per damunt de la mitjana del 59,55%, mentre que el mes que registra menys pernoctacions és gener, el qual registra tan sols uns valors que corresponen al 58,95% del valor de referència *M*. Podem observar també que els valors més importants es concentren als mesos d’agost, juliol i setembre, mentre que els mesos de menor afluència corresponen als mesos de gener, desembre i febrer.

Podem observar-hi un comportament estacional molt marcat.

Taula corresponent a la sèrie de pernoctacions a la Comunitat Valenciana (2006-2010)

	2006	2007	2008	2009	2010		$\overline{y}_k$	$\overline{y}'_k$	e_k	I_k
Gener	1296648	1307384	1289627	1115865	1061450		1214195	1214195	–845348	58,95
Febrer	1424453	1472806	1562471	1329013	1347885		1427326	1430862	–628681	69,47
Març	1750282	1892925	1993175	1700733	1800323		1827488	1834560	–224982	89,08
Abril	2152783	2226734	1868049	1925454	1943080		2023220	2033828,6	–25714	98,75
Maig	2131194	2181889	2114631	1952409	2049616		2085948	2100092,6	40550	101,97
Juny	2399782	2523555	2311444	2241147	2049616		2305109	2322789,8	263247	112,78
Juliol	2884491	2983227	2877028	2871011	2945899		2912331	2933548,4	874006	142,44
Agost	3153407	3308833	3227055	3274561	3341961		3261163	3285916,8	1226374	159,55
Setembre	2540711	2580090	2524888	2381776	2424567		2490406	2518696	459154	122,29
Octubre	2207203	2097005	2012838	1954615	2053500		2065032	2096858	37316	101,81
Novembre	1674433	1741296	1498967	1471579	1472610		1571777	1607139	–452403	78,03
Desembre	1437036	1420988	1251809	1201342	1174451	**TOTALS**	1297125	1336023,4	–723519	64,87
$\overline{y}_i$	2087702	2144728	2044332	1951625	1972080	10200467		**M** = 2059542		
i	–2	–1	0	1	2	0				
$\overline{y}_i \cdot i$	–4175404	–2144728	0	1951625	3944160	-424346				
i^2	4	1	0	1	4	10				
Tendència	2124962,6	2082528	2040093	1997658,8	1955224,2		b/m=-42434,6/12=-3536,2			

Recta de tendència: $T_i = a + bi = 2040093,4 - 42434,6i$ amb $i = 0$ que correspon a l'any 2008 com podem veure a la taula.



Càlcul de la component residual o erràtica

Per a calcular la component residual $r_{ik} = y_{ik} - T_i - e_k$ caldrà determinar primerament el valor de la tendència per a cada any i considerarem per a la component estacional e_k els valors que podem trobar en la taula ja calculats. Així:

$$T_{2006} = 2040093,4 - 42434,6(-2) = 2124962,6$$

$T_{2007} = 2040093{,}4 - 42434{,}6(-1) = 2082528$
$T_{2008} = 2040093{,}4 - 42434{,}6 \cdot 0 = 2040093{,}4$
$T_{2009} = 2040093{,}4 - 42434{,}6 \cdot 1 = 1997658{,}8$
$T_{2010} = 2040093{,}4 - 42434{,}6 \cdot 2 = 1955224{,}2$

i considerarem els valors de la component estacional:

$e_1 = -845.348$	$e_2 = -628.681$	$e_3 = -224.982$
$e_4 = -25.714$	$e_5 = 40.550$	$e_6 = 263.247$
$e_7 = 874.00$	$e_8 = 1.226.374$	$e_9 = 459.154$
$e_{10} = 37.316$	$e_{11} = -452.403$	$e_{12} = -723.519$

Amb aquestes dades, els valors de la component residual estan calculats en la taula següent (per veure millor les dades a efectes d'interpretació hem arrodonit els resultats a nombre exacte de pernoctacions i hem ressenyat amb roig els valors negatius).

Per ser la component residual, aquestes xifres podrien sorprendre'ns, però hem de considerar que si comparem els valors més extrems ressaltats (que posteriorment comentarem), estem parlant de magnituds properes a +/– 150000 front al valor global de la taula $M = 2059542$. Per la qual cosa, estem parlant de valors relatius propers al 7%.

Fem aquest advertiment perquè tots sabem que les quantitats que cal obtenir de la component residual haurien de ser menudes en valor absolut, i que no presenten cap regularitat. Ja sabem que estem calculant les quantitats no explicades pel nostre model i que permetran ressaltar aquells valors puntuals, que per raons no predictibles, mostren divergència del valor que caldria esperar, atenent les components de la tendència i estacional.

Com es pot veure, al març i agost del 2006, els valors van estar sota les previsions, mentre que als mesos d'abril i juny del 2007 ho van fer per damunt. Cal fixar-se amb el comportament de les dades als mesos de febrer, març i abril del 2008 i juny, juliol i agost del 2010 que presenten en ambdós casos la seqüència de tres mesos amb valors que sembla que es «compensen» en la mateixa temporada.

	residual 2006	**residual 2007**	**residual 2008**	**residual 2009**	**residual 2010**
Gener	17033	70204	94881	-36446	-48427
Febrer	-71829	18959	**151058**	-39965	21341
Març	**-149698**	35379	**178064**	-71943	70081
Abril	53534	**169920**	**-146331**	-46491	13570
Maig	-34319	58811	33987	-85800	53842
Juny	11572	**177780**	8103	-19759	**-168856**

Juliol	-114478	26693	-37071	-654	**116669**
Agost	**-197930**	-69	-39413	50528	**160362**
Setembre	-43405	38408	25641	-75036	10189
Octubre	44925	-22839	-64571	-80359	60960
Novembre	1874	111171	-88723	-73676	-30211
Desembre	35592	61979	-64765	-72798	-57254

Previsions per a l'any 2012 i 2013

Per a fer prediccions per als anys propers recordem què implica que l'anàlisi del nostre model siga vigent i que cap altra circumstància aliena altere les regularitats que hem ressenyat amb el nostre model (la tendència lleugerament decreixent i el comportament mensual ja comentat).

Si volem preveure les quantitats de l'any 2012 i 2013 caldrà calcular la tendència substituint els valors $i = 4$ (per a l'any 2012) i $i = 5$ (per a l'any 2013). Aquests valors serien els que correspondrien a aquests anys en l'escala dels valors i de la primera part de la taula. Així:

$T_{2012} = 2040093{,}4 - 42434{,}6 \cdot 4 = 1870355$
$T_{2013} = 2040093{,}4 - 42434{,}6 \cdot 5 = 1827920{,}4$

i amb la component estacional abans esmentada, podrem fer les previsions mensuals per als anys 2012 i 2013, calculant $y_{ik} = T_i + e_k$.

	Previsions 2012	**Previsions 2013**
Gener	1870355 – 845348 = 1025007	1827920,4 – 845348 = 982573
Febrer	1870355 – 628681 = 1241674	1827920,4 – 628681 = 1199240
Març	1870355 – 224982 = 1645373	1827920,4 – 224982 = 1602938
Abril	1870355 – 25714 = 1844641	1827920,4 – 25714 = 1802207
Maig	1870355 + 40550 = 1910905	1827920,4 + 40550 = 1868471
Juny	1870355 + 263247 = 2133602	1827920,4 + 263247 = 2091168
Juliol	1870355 + 874006 = 2744361	1827920,4 + 874006 = 2701926
Agost	1870355 + 1226374 = 3096729	1827920,4 + 1226374 = 3054295
Setembre	1870355 + 459154 = 2329509	1827920,4 + 459154 = 2287074
Octubre	1870355 + 37316 = 1907671	1827920,4 + 37316 = 1865236
Novembre	1870355 – 452403 = 1417952	1827920,4 – 452403 = 1375517
Desembre	1870355 – 723519 = 1146836	1827920,4 – 723519 = 1104401

Exercici 3

Amb les dades següents, extretes de la DGT, que ens mostren les noves llicències de tots els tipus de carnets de conduir a la Comunitat Valenciana, des de l'any 2008 fins al 2010, realitzeu una anàlisi del fenomen, obtenint les diferents components de la sèrie temporal per l'ajust analític.

Interpreteu el significat de cadascuna de les components.

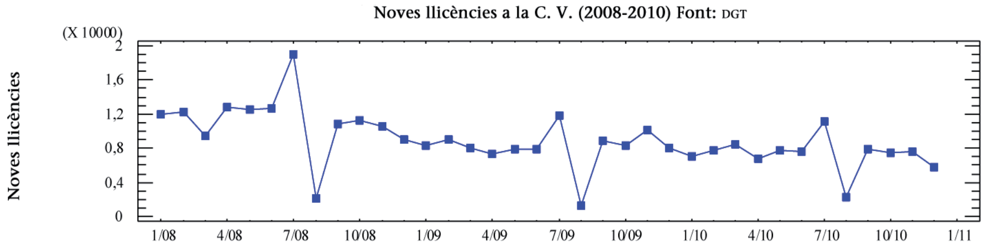
Feu les previsions que podem esperar per els anys 2011 i 2012.

	2008	2009	2010
Gener	12031	8380	7071
Febrer	12208	8993	7685
Març	9497	7973	8444
Abril	12862	7360	6781
Maig	12567	7874	7728
Juny	12723	7881	7585
Juliol	19003	11820	11138
Agost	2147	1346	2205
Setembre	10826	8876	7901
Octubre	11196	8374	7427
Novembre	10628	10137	7665
Desembre	9064	8083	5746

Solució

Anàlisi del model i gràfica

Per començar, cal observar la gràfica de les dades i comprovar que poden ser analitzades pel mètode de l'ajust analític, suposant un model additiu.



Podem observar una tendència lleugerament decreixent i un comportament estacional prou constant, exceptuant els mesos de juliol i agost.

Càlcul de la tendència

Considerarem les dades de la sèrie següent que fa referència al nombre de les noves llicències de la totalitat dels tipus de permisos de conduir en la Comunitat Valenciana, en els anys 2008 al 2010:

	2008	2009	2010	
Gener	12031	8380	7071	
Febrer	12208	8993	7685	
Març	9497	7973	8444	
Abril	12862	7360	6781	
Maig	12567	7874	7728	
Juny	12723	7881	7585	
Juliol	19003	11820	11138	
Agost	2147	1346	2205	
Setembre	10826	8876	7901	
Octubre	11196	8374	7427	
Novembre	10628	10137	7665	
Desembre	9064	8083	5746	TOTALS
$\overline{y}_i$	11229	8091	7281	26602
i	-1	0	1	0
$\overline{y}_i \cdot i$	-11229	0	7281	-3948
i^2	1	0	1	2

Podem veure que a la part inferior de la taula hem calculat les mitjanes mensuals de cada any $\overline{y}_i$, els valors de l'escala i, i les dues files inferiors $\overline{y}_i \cdot i$, i^2, que ens permetrà calcular en la columna TOTALS la suma dels valors de cada fila.

Amb aquestes dades, plantejarem el sistema següent:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases} \Rightarrow \begin{cases} 26602 = 3a + 0b \\ \text{-}3948 = 0a + 2b \end{cases}$$

on a i b són els coeficients de la recta de regressió que denominem recta de tendència, la fórmula de la qual és $T_i = a + b$, on i fa referència a l'any que se indica en l'escala de la taula superior.

Podem veure en aquest exemple que l'estratègia de donar-li el valor $i = 0$ a l'any corresponent a la columna central, quan N (nombre d'anys) és imparell simplifica molt la resolució del sistema. Així:

$$a = \frac{26602}{3} = 8867{,}36 \qquad\qquad b = \frac{-\ 3948}{2} = -\ 1974$$

són els coeficients de $T_i = a + bi = 8867{,}36 - 1974i$ que és l'equació de la recta de tendència i ens permet interpretar que a llarg termini, a partir d'aquestes dades, la mitjana mensual del nombre de noves llicències disminueix en 1974 nous permisos per any. Per tractar-se d'una quantitat menuda en relació a les dades de la sèrie és la raó per la qual a la gràfica no hi observàvem clarament el decreixement.

Càlcul de la component estacional

Per a calcular la component estacional, treballarem les columnes que es poden veure a la dreta de la taula següent.

Recordem que en la primera columna, $\overline{y}_k = \frac{\sum_i y_{ij}}{N}$ són les mitjanes de les dades de cada mes, calculades amb les dades originals, per files.

La columna següent correspon a les mitjanes corregides, la fórmula de les quals és $\overline{y}'_k = \overline{y}_k - \frac{b}{m}(k - 1)$ on $b/m = -\ 1974/12 = -164{,}5$. Recordem que $m = 12$ perquè, en tractar-se de dades mensuals, en tenim 12 per any.

M és la mitjana global corregida i és la mitjana de les mitjanes corregides.

$$M = \frac{\sum_k \overline{y}'_k}{m}$$

que podem interpretar com el valor mitjà de les noves mitjanes corregides mensuals i que representarà el 100% o valor de referència.

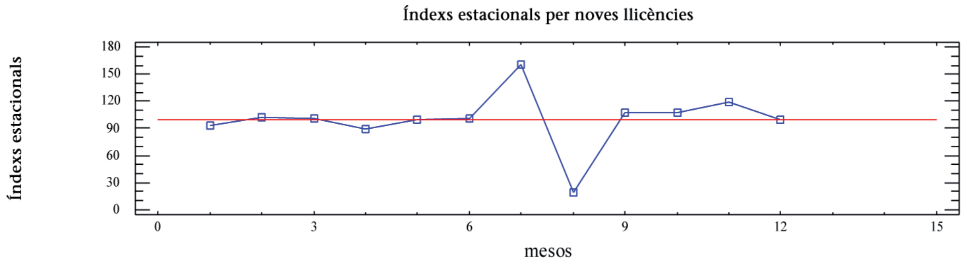
Després, les dues columnes següents mostren els valors de la component estacional $e_k = \bar{y}_k' - M$ i els índexs estacionals $I_k = \frac{\bar{y}_k}{M} \cdot 100$ on podem veure en quins mesos el nombre de noves llicències està per damunt o sota el valor de M. e_k ens presenta aquesta desviació en quantitats absolutes (nombre de pernoctacions) mentre que I_k ho indica de manera percentual.

Podem interpretar que el mes amb menys llicències noves expedides és agost amb un valor molt per davall de la mitjana del 31%, mentre que el mes que registra més llicències expedides és juliol que registra un valor del 53% per damunt del valor de referència *M*. Podem observar també que la resta de mesos tenen valors pròxims al 100%, la qual cosa es pot interpretar com que aquest fenomen no té un comportament estacional marcat, tret dels mesos de juliol i agost abans citats.

Taula corresponent a la sèrie de noves llicències expedides de tots els tipus de carnets a la Comunitat Valenciana (2008-2010)

	2008	2009	2010			$\bar{y}_k$	$\bar{y}_k'$	e_k	I_k
Gener	12031	8380	7071			9161	9.161	–611	93,74
Febrer	12208	8993	7685			9629	9.793	21	100,22
Març	9497	7973	8444			8638	8967	–805	91,76
Abril	12862	7360	6781			9001	9494,5	–278	97,16
Maig	12567	7874	7728			9390	10047,6667	276	102,82
Juny	12723	7881	7585			9396	10218,8333	447	104,57
Juliol	19003	11820	11138			13987	14974	5202	153,23
Agost	2147	1346	2205			1899	3050,83333	–6721	31,22
Setembre	10826	8876	7901			9201	10517	745	107,62
Octubre	11196	8374	7427			8999	10479,5	707	107,24
Novembre	10628	10137	7665			9477	11121,6667	1350	113,81
Desembre	9064	8083	5746	TOTALS		7631	9440,5	–332	96,61
$\bar{y}_i$	11229	8091	7281	26602		M =	9.772		
i	–1	0	1	0					
$\bar{y}_i \times i$	–11229	0	7281	–3948					
i^2	1	0	1	2					
Tendència	10841,36111	8867,361	6893,361111						
					b/m = –164,5				

Recta de tendència: $T_i = a + bi = 8867,36 - 1974i$ amb $i = 0$ que correspon a l'any 2009 com podem veure en la taula.



Càlcul de la component residual o erràtica

Per a calcular la component residual $r_{ik} = y_{ik} - T_i - e_k$ caldrà calcular primerament el valor de la tendència per a cada any i considerarem per a la component estacional e_k els valors que podem trobar en la taula ja calculats. Així:

$T_{2008} = 8867,36 – 1974(–1) = 10841,36$
$T_{2009} = 8867,36 – 1974 \cdot 0 = 8867,36$
$T_{2010} = 8867,36 – 1974 \cdot 1 = 6893,36$

i considerarem els valors de la component estacional:

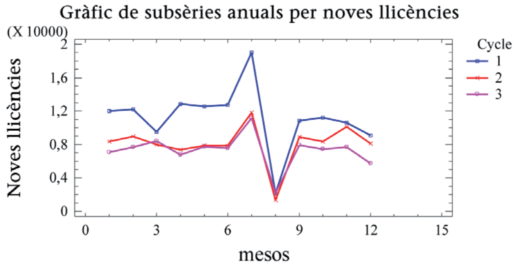
e_1 = -611	e_2 = 21	e_3 = -805
e_4 = -278	e_5 = 276	e_6 = 447
e_7 = 5.202	e_8 = -6.721	e_9 = 745
e_{10} = 707	e_{11} = 1.350	e_{12} = -332

Amb aquestes dades, els valors de la component residual estan calculats en la taula següent (per veure millor les dades a efectes d'interpretació hem arrodonit els resultats de llicències i hem ressenyat amb roig els valors negatius).

Fem aquest advertiment perquè tots sabem que les quantitats que cal obtenir de la component residual haurien de ser menudes en valor absolut, i que no presenten cap regularitat. Ja sabem que estem calculant les quantitats no explicades pel nostre model i que permetran ressaltar aquells valors puntuals, que per raons no predictibles, mostren divergència del valor que caldria esperar, atenent les components de la tendència i estacional.

Com es pot veure, en la taula següent hem remarcat en roig les dades que estan per davall dels valors que caldria esperar i amb blau els valors que estan per damunt. Per a interpretar conjuntament les dades reals i els resultats de la component estacional hem introduït davall la taula un gràfic de les subsèries anuals de les dades.

	residual 2008	residual 2009	residual 2010
Gener	1801	124	789
Febrer	1346	105	771
Març	–539	–89	2356
Abril	2298	–1230	165
Maig	1450	–1269	559
Juny	1435	–1433	245
Juliol	2960	–2249	–957
Agost	–1973	–800	2033
Setembre	–760	–736	263
Octubre	–353	–1201	–174
Novembre	–1563	–80	–578
Desembre	–1446	–453	–816



Podem observar que els valors de les dades de la component residual de l'any 2008, efectivament, corresponen als mesos, les dades dels quals s'allunyen prou del comportament que correspon al patró dels altres anys. Les dades d'agost distorsionen un poc perquè no segueixen el comportament de la tendència anual, és com si diguérem que és un mínim perquè paren els serveis i és independent dels valors de la tendència.

La major part de les dades ressenyades corresponen a l'any 2008 perquè es pot observar millor en aquest gràfic que no segueix el mateix patró estacional que els anys següents, si exceptuem els valors dels mesos d'estiu.

En tot cas, cal destacar que és una sèrie massa curta (tres anys) per a fer una anàlisi molt rigorosa.

Previsions per a l'any 2012 i 2013

Recordem que fer prediccions per als anys propers implica que l'anàlisi del nostre model siga vigent i que cap altra circumstància aliena altere les regularitats que hem ressenyat amb el nostre model (la tendència lleugerament decreixent i el comportament mensual ja comentat).

Si volem preveure les quantitats de l'any 2012 i 2013 caldrà calcular la tendència substituint els valors $i = 2$ (per a l'any 2011) i $i = 3$ (per a l'any 2012). Aquests valors serien els que correspondrien a aquests anys en l'escala dels valors i de la primera part de la taula. Així:

$$T_{2011} = 8867{,}36 - 1974 \cdot 2 = 4919{,}36$$

$$T_{2012} = 8867{,}36 - 1974 \cdot 3 = 2945{,}36$$

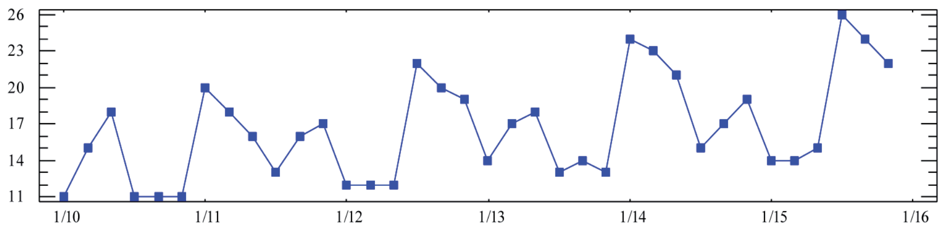
i amb la component estacional abans esmentada, podrem fer les previsions mensuals per als anys 2011 i 2012, calculant $y_{ik} = T_i + e_k$.

	Previsions 2011	Previsions 2012
Gener	4919,36 – 611 = 4308	2945,36 – 611 = 2334
Febrer	4919,36 + 21 = 4940	2945,36 + 21 = 2966
Març	4919,36 – 805 = 4114	2945,36 – 805 = 2140
Abril	4919,36 – 278 = 4642	2945,36 – 278 = 2668
Maig	4919,36 + 276 = 5195	2945,36 + 276 = 3221
Juny	4919,36 + 447 = 5366	2945,36 + 447 = 3292
Juliol	4919,36 + 5202 = 10.121	52945,36 + 202 = 8147
Agost	4919,36 – 6721 = –1802	2945,36 – 6.721 = –3776
Setembre	4919,36 + 745 = 5664	2945,36 + 745 = 3690
Octubre	4919,36 + 707 = 5627	2945,36 + 707 = 3653
Novembre	4919,36 + 1350 = 6269	2945,36 + 1.350 = 4295
Desembre	4919,36 – 332 = 4588	2945,36 – 332 = 2614

Cal fer notar que els resultats de la taula anterior estan arrodonits per tractar-se del nombre de noves llicències que podem preveure per als anys que s'assenyalen. No s'ha d'oblidar, insistim, que la fiabilitat d'aquestes previsions està en funció de la hipòtesi que el fenomen mantinga el comportament decreixent que ens ha indicat la tendència.

Exercici 4

La gràfica següent és la representació d'una sèrie temporal on es detallen les dades bimensuals de 6 anys. Si haguérem de calcular la tendència i la component estacional de la sèrie pel mètode de les mitjanes mòbils, explica l'elecció del nombre de dades que cal considerar per al càlcul de les mitjanes (p) en cada cas, i justifica la resposta.



Per a començar la resolució pel mètode de les mitjanes mòbils cal triar el nombre adient d'observacions a incloure en cada mitjana. Aquest nombre depén de la component que volem trobar.

Solució

Començarem per explicar aquest nombre en el cas de la tendència.

Aquest nombre que denotarem per p és el mínim comú múltiple, MCM d'ara endavant, del nombre d'observacions per any, (en aquest cas 6 per ser observacions bimensuals, és a dir, tenim una dada cada dos mesos, per la qual cosa en tenim 6 per any) i el nombre d'observacions que inclou cada «període» de la gràfica (9 en aquest cas perquè tal com s'hi veu, la sèrie té un comportament o patró que es repeteix cada 9 punts aproximadament).

Així doncs, per a calcular la tendència el nombre d'observacions a prendre per a cada mitjana mòbil es p, on

$$p = MCM(6,9) = 18$$

Aquesta tria ens assegura que en fer el suavitzat de 18 observacions, totes les fluctuacions anuals i estacionals es contemplen en cada mitjana, i així substituïm cada valor que s'abandona en una mitjana per un altre que seria equivalent a la mitjana següent per tal d'aconseguir la tendència, que explica el comportament del fenomen a llarg termini de la sèrie.

Estudiem ara quin seria el nombre d'observacions a considerar en el cas de la component estacional.

En aquest cas, *p* és el nombre d'observacions que cal agafar per tenir un any complet. És a dir, en el nostre problema, $p = 6$.

En aquesta tria, cada dada és substituïda en una mitjana per una altra que té el mateix comportament estacional l'any vinent en la mitjana següent. Podrem observar que en cadascun dels càlculs sempre considerarem les dades d'un any complet.

Exercici 5

En la taula següent presentem el nombre total de viatgers transportats en els serveis de transport públic a la Comunitat Valenciana, detallats per mesos, dels anys 2006 al 2010.

	2006	2007	2008	2009	2010
Gener	12340	12542	12622	11869	10283
Febrer	11850	12156	12202	11365	10981
Març	13721	13729	10457	12024	12207
Abril	10919	11612	12960	10469	10549
Maig	13495	13777	12713	12038	12076
Juny	13029	13094	12619	12314	11783
Juliol	12118	12364	12351	11490	10621
Agost	8803	8814	8846	8027	7698
Setembre	12148	11768	12057	10964	10705
Octubre	13141	13266	13277	11956	11447
Novembre	13307	12655	12474	11718	11557
Desembre	11859	11624	11682	10894	10769

Solució

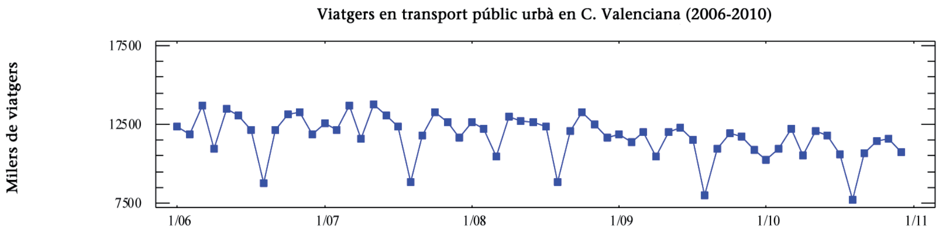
a) Calculeu les components d'aquesta sèrie, pel mètode de les mitjanes mòbils. Interpreteu-ne els resultats.

Anàlisi de les dades per a decidir el nombre d'observacions que cal prendre en cada mitjana

Per a començar la resolució pel mètode de les mitjanes mòbils cal triar el nombre adient d'observacions a incloure en cada mitjana. Aquest nombre depén de la component que vulguem trobar.

Començarem per explicar aquest nombre en el cas de la tendència.

Aquest nombre que denotarem per *n* és el mínim comú múltiple, MCM d'ara endavant, del nombre d'observacions per any, (en aquest cas 12 per ser observacions mensuals) i el nombre d'observacions que inclou cada «període» de la gràfica (12 en aquest cas perquè tal com s'hi veu, la gràfica següent té un comportament que es repeteix cada any).



Per a calcular la tendència

Així, cal fer les mitjanes aritmètiques de 12 dades, de manera que en el primer cas, calculem la seua mitjana, que seria:

$$y_{1-12} = \frac{12340+11850+13721+10919+...+12148+13141+13307+11859}{12} = 12227{,}5$$

i així anem substituint una dada (la primera) per la següent de la sèrie, de tal manera que sempre fem la mitjana de 12 dades, és a dir, calculem pel mètode de les mitjanes mòbils de 12 dades la segona columna que utilitzarem per a calcular la tendència de la sèrie. Es pot observar que sempre es tracta de la mitjana de les dades de tot un any, encara que aquesta col·lecció comença en cada mitjana pels diferents mesos:

Podem observar com hem calculat els següents termes de la segona columna:

$$y_{2-13} = \frac{11850+13721+10919+...+13141+13307+11859+12542}{12} = 12244{,}3333$$

$$y_{3-14} = \frac{13721+10919+13495+...+13307+11859+12542+12156}{12} = 12269{,}8333$$

$$y_{4-15} = \frac{10919+13495+13029+...+11859+12542+12156+13729}{12} = 12270{,}5$$

$$y_{5-16} = \frac{113495+13029+12118+...+11859+12542+12156+13729}{12} = 12328{,}25$$

i així, successivament, es calculen els valors restants de la segona columna que hem denotat per $\bar{y}$.

Un tema que cal especificar és que aquestes mitjanes corresponen a 12 mesos (nombre parell) per la qual cosa és difícil fer-les correspondre a cap dels períodes dels quals partim, ja que no tindrem un període que corresponga al centre de les observacions promediades. En aquests casos en què la mitjana correspon a un nombre parell de dades, s'ha de calcular la tercera columna on hi haurà les mitjanes mòbils centrades i, a més, és la mitjana aritmètica de cada dues dades consecutives de la segona columna, per a fer-les correspondre a un mes en particular.

Expliquem com obtenir els primers valors de la tercera columna, denotada per T_{ij}, els valors de la qual considerem que expliquen la «tendència» de la sèrie i ens permet veure el comportament dels valors de la sèrie a llarg termini. A diferència del mètode de l'ajust analític, la tendència és una seqüència numèrica extreta dels valors de les dades i no disposem d'una expressió algebraica per a obtenir un valor de tendència per a cada any.

En aquest mètode de les mitjanes mòbils obtenim un valor diferent que fa referència a cada mes de cada any, exceptuant els valors inicials i finals que no podem calcular.

Gener 2006	12340		
Febrer 2006	11850		
Març 2006	13721		
Abril 2006	10919		
Maig 2006	13495		
Juny 2006	13029	12227,5	
Juliol 2006	12118	12244,3333	12235,9167
Agost 2006	8803	12269,8333	12257,0833
Setembre 2006	12148	12270,5	12270,1667
Octubre 2006	13141	12328,25	12299,375
Novembre 2006	13307	12351,75	12340
Desembre 2006	11859	12357,1667	12354,4583
Gener 2007	12542	12377,6667	12367,4167
Febrer 2007	12156	12378,5833	12378,125
Març 2007	13729	12346,9167	12362,75
Abril 2007	11612	12357,3333	12352,125
Maig 2007	13777	12303	12330,1667
Juny 2007	13094	12283,4167	12293,2083
Juliol 2007	12364	12290,0833	12286,75

Agost 2007	8814	12293,9167	12292
Setembre 2007	11768	12021,25	12157,5833
Octubre 2007	13266	12133,5833	12077,4167
Novembre 2007	12655	12044,9167	12089,25
Desembre 2007	11624	12005,3333	12025,125
Gener 2008	12622	12004,25	12004,7917
Febrer 2008	12202	12006,9167	12005,5833
Març 2008	10457	12031	12018,9583
Abril 2008	12960	12031,9167	12031,4583
Maig 2008	12713	12016,8333	12024,375
Juny 2008	12619	12021,6667	12019,25
Juliol 2008	12351	11958,9167	11990,2917
Agost 2008	8846	11889,1667	11924,0417
Setembre 2008	12057	12019,75	11954,4583
Octubre 2008	13277	11812,1667	11915,9583
Novembre 2008	12474	11755,9167	11784,0417
Desembre 2008	11682	11730,5	11743,2083
Gener 2009	11869	11658,75	11694,625
Febrer 2009	11365	11590,5	11624,625
Març 2009	12024	11499,4167	11544,9583
Abril 2009	10469	11389,3333	11444,375
Maig 2009	12038	11326,3333	11357,8333
Juny 2009	12314	11260,6667	11293,5
Juliol 2009	11490	11128,5	11194,5833
Agost 2009	8027	11096,5	11112,5
Setembre 2009	10964	11111,75	11104,125
Octubre2009	11956	11118,4167	11115,0833
Novembre 2009	11718	11121,5833	11120
Desembre 2009	10894	11077,3333	11099,4583
Gener 2010	10283	11004,9167	11041,125
Febrer 2010	10981	10977,5	10991,2083
Març 2010	12207	10955,9167	10966,7083
Abril 2010	10549	10913,5	10934,7083
Maig 2010	12076	10900,0833	10906,7917
Juny 2010	11783	10889,6667	10894,875

Juliol 2010	10621		
Agost 2010	7698		
Setembre 2010	10705		
Octubre 2010	11447		
Novembre 2010	11557		
Desembre 2010	10769		

Podem presentar els càlculs dels primers valors de la tendència:

$$T_{juliol\,2006} = \frac{12227{,}5 + 12244{,}3333}{2} = 12235{,}9167$$

$$T_{agost\,2006} = \frac{12244{,}3333 + 12269{,}8333}{2} = 12257{,}0833$$

$$T_{setembre\,2006} = \frac{12269{,}8333 + 12270{,}5}{2} = 12270{,}1667$$

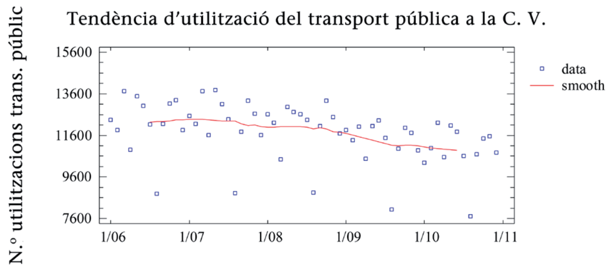
i així podem comprovar els valors de la resta de la columna.

Caldria explicar millor que els valors de la segona i tercera columna, potser quedarien millor referits al període, si les presentàrem com ho fem en la taula següent. Hem calculat les mitjanes de 12 valors i les fem correspondre al «centre» d'aquest període, que situem en la cel·la de la taula que hem previst intercalant una fila entre cada dues files de dades originals, i que després permetran veure més clarament les dades d'aquest procediment de mitjanes mòbils centrades. No hem presentat els càlculs de tota la sèrie, per motius d'espai, com és evident. No obstant això, sí que volem mostrar-ho amb els primers valors de la sèrie:

	y_k		T_k
Gener 2006	12340		
Febrer 2006	11850		
Març 2006	13721		
Abril 2006	10919		
Maig 2006	13495		
Juny 2006	13029		
		12227,5	
Juliol 2006	12118		12235,9167
		12244,3333	

Agost 2006	8803		12257,0833
		12269,8333	
Setembre 2006	12148		12270,1667
		12270,5	
Octubre 2006	13141		12299,375
		12328,25	
Novembre 2006	13307		12340
		12351,75	
Desembre 2006	11859		12354,4583
		12357,1667	
Gener 2007	12542		12367,4167
		12377,6667	
Febrer 2007	12156		12378,125
		12378,5833	
Març 2007	13729		12362,75
		12346,9167	
Abril 2007	11612		12352,125
		12357,3333	
Maig 2007	13777		12330,1667
		12303	
Juny 2007	13094		12293,2083
		12283,4167	
Juliol 2007	12364		12286,75
		12290,0833	
Agost 2007	8814		12292
		12293,9167	
Setembre 2007	11768		12157,5833
		12021,25	
Octubre 2007	13266		12077,4167
		12133,5833	
Novembre 2007	12655		12089,25
		12044,9167	
Desembre 2007	11624		12025,125

Es pot observar que les primeres i darreres dades no tenen les seues mitjanes corresponents per la tècnica utilitzada. Podrem interpretar millor el sentit d'aquestes dades calculades amb la representació dels valors damunt de la sèrie, que ens permet interpretar el comportament a llarg termini dels valors de la utilització del transport públic a la Comunitat Valenciana:



Podem veure que a llarg termini, si considerem les dades que hi analitzem, la utilització del transport públic a la Comunitat Valenciana sembla disminuir. Té un comportament decreixent si observem al gràfic anterior la línia roja que representa els valors de la tendència.

Per a calcular la component estacional

Per a continuar l'anàlisi s'ha de calcular ara la component estacional. A tal fi, cal que determinem els valors de les mitjanes aritmètiques dels valors de cada mes. Amb aquest propòsit, situarem les dades originals distribuïdes en files per períodes i en columnes per anys, com es pot veure a la taula següent. En la columna de la dreta hem calculat les mitjanes ja citades,

$$\bar{y}_k = \frac{\sum_i y_{ij}}{N}.$$

	2006	2007	2008	2009	2010	$\bar{y}_k$
Gener	12340	12542	12622	11869	10283	11931,2
Febrer	11850	12156	12202	11365	10981	11710,8
Març	13721	13729	10457	12024	12207	12427,6
Abril	10919	11612	12960	10469	10549	11301,8
Maig	13495	13777	12713	12038	12076	12819,8
Juny	13029	13094	12619	12314	11783	12567,8
Juliol	12118	12364	12351	11490	10621	11788,8
Agost	8803	8814	8846	8027	7698	8437,6
Setembre	12148	11768	12057	10964	10705	11528,4
Octubre	13141	13266	13277	11956	11447	12617,4
Novembre	13307	12655	12474	11718	11557	12342,2
Desembre	11859	11624	11682	10894	10769	11365,6

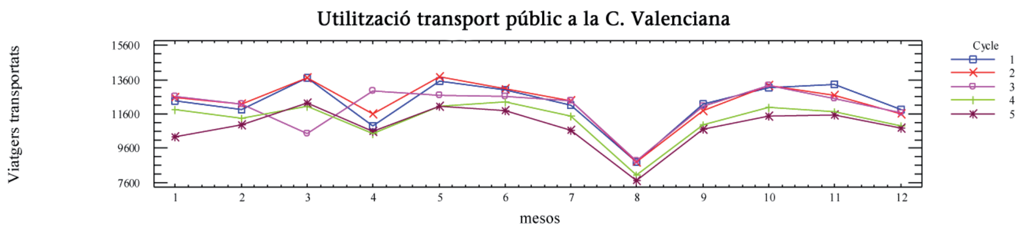
Per altra banda, situarem en la taula següent, els valors de les mitjanes mòbils que s'han de calcular amb un nombre de dades p que és el nombre de dades que tenim per any. En aquest problema, podrem aprofitar els valors de la tercera columna del càlcul de la tendència, ja que en aquest cas també vam utilitzar mitjanes mòbils amb $p = 12$. Així doncs, redistribuirem els valors d'aquestes mitjanes en una taula on cada cel·la indicarà el valor que correspon a cada mes i any.

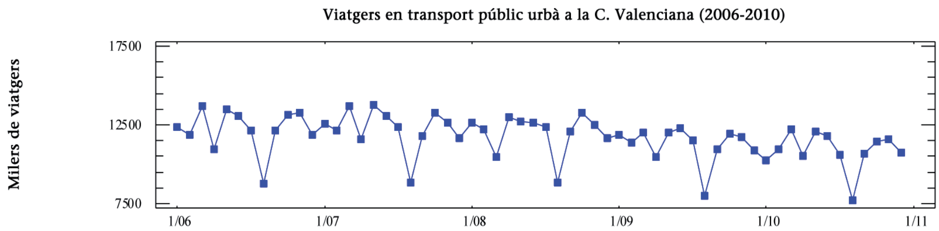
	2006	2007	2008	2009	2010	E_k
Gener		12367,4167	12004,7917	11694,625	11041,125	11776,9896
Febrer		12378,125	12005,5833	11624,625	10991,2083	11749,8854
Març		12362,75	12018,9583	11544,9583	10966,7083	11723,3437
Abril		12352,125	12031,4583	11444,375	10934,7083	11690,6667
Maig		12330,1667	12024,375	11357,8333	10906,7917	11654,7917
Juny		12293,2083	12019,25	11293,5	10894,875	11625,2083
Juliol	12235,9167	12286,75	11990,2917	11194,5833		11926,8854
Agost	12257,0833	12292	11924,0417	11112,5		11813,8568
Setembre	12270,1667	12157,5833	11954,4583	11104,125		11757,5058
Octubre	12299,375	12077,4167	11915,9583	11115,0833		11716,491
Novembre	12340	12089,25	11784,0417	11120		11677,4457
Desembre	12354,4583	12025,125	11743,2083	11099,4583		11636,3093

Per a calcular la component estacional caldrà restar les columnes finals d'aquestes dues taules, $e_k = \bar{y}_k - E_k$, cosa que ens permetrà explicar quins són els mesos de major i menor utilització del transport públic a la Comunitat Valenciana.

	e_k
Gener	154,2104
Febrer	–39,0854
Març	704,256275
Abril	–388,86665
Maig	1165,00833
Juny	942,591675
Juliol	–138,085425
Agost	–3376,25678
Setembre	–229,105845
Octubre	900,908964
Novembre	664,754316
Desembre	–270,709321

Així doncs, vegem els resultats de la taula anterior, on podem destacar que el mes de major utilització del transport públic a la Comunitat Valenciana es produeix al mes de maig, seguit també del mesos de juny i octubre, pel contrari es veu que el mes que destaca perquè el seu valor és molt inferior a la mitjana és agost. Aquesta interpretació podem comprovar-la perquè els màxims i els mínims de la gràfica de la sèrie coincideixen amb aquests valors de la component estacional.





Per a calcular la component residual o erràtica

Per a calcular la component erràtica o residual cal restar a cada valor inicial el valor de la tendència corresponent a cada cel·la en els casos en què hi existeix (no en els primers i darrers valors de la sèrie) i també li restem el valor de la component estacional que correspon a cada període.

En la taula següent indiquem aquests càlculs. Així, la component erràtica per a cada període serà:

	2006	2007	2008	2009	2010
Gener		12542-12367,4167--154,2104	12622-12004,7917 -154,2104	11869-11694,625 -154,2104	10283 -11041,125 -154,210
Febrer		12156-12378,125 +39,0854	12202-12005,5833 +39,0854	11365-11624,625 +39,0854	10981-10991,2083 +39,0854
Març		13729-12362,75 -704,256275	10457-12018,9583 -704,256275	12024-11544,9583 -704,256275	12207-10966,7083 -704,256275
Abril		11612-12352,125 +388,86665	12960-12018,9583 +388,86665	10469-11444,375 +388,86665	10549-10934,7083 +388,86665
Maig		13777-12330,1667 -1165,00833	12713-12024,375 -1165,00833	12038-11357,8333 -1165,00833	12076-10906,7917 -1165,00833
Juny		13094-12293,2083 -942,591675	12619-12019,25 -942,591675	12314-11293,5 -942,591675	11783-10894,875 -942,591675
Juliol	12118 - 12235,9167 +138,085425	12364-12286,75 +138,085425	12351-11990,2917 +138,085425	11490-11194,5833 +138,085425	
Agost	8803 - 12257,0833 +3376,25678	8814-12292 +3376,25678	8846-11924,0417 +3376,25678	8027 - 11112,5 +3376,25678	
Setembre	12148 - 12270,1667 +229,105845	11768-12157,5833 +229,105845	12057-11954,4583 +229,105845	10964-11104,125 +229,105845	
Octubre	13141-12299,375 -900,908964	13266-12077,4167 -900,908964	13277-11915,9583 -900,908964	11956-11115,0833 -900,908964	
Novembre	13307 – 12340 - 664,754316	12655-12089,25 -664,754316	12474-11784,0417 -664,754316	11718-11120 -664,754316	
Desembre	11859-12354,4583 +270,709321	11624-12025,125 +270,709321	11682-11743,2083 +270,709321	10894-11099,4583 +270,709321	

Per tant, podem concloure que la variable residual o erràtica corresponent a cada període són els valors de la taula següent:

	2006	2007	2008	2009	2010
Gener		20,3729	462,9979	20,1646	-912,335
Febrer		-183,0396	235,5021	-220,5396	28,8771
Març		661,993725	-2266,21458	-225,214575	536,035425
Abril		-351,25835	1329,90835	-586,50835	3,15835
Maig		281,82497	-476,38333	-484,84163	4,19997
Juny		-141,799975	-342,841675	77,908325	-54,466675
Juliol	20,168725	215,335425	498,793725	433,502125	
Agost	-77,82652	-101,74322	298,21508	290,75678	
Setembre	106,939145	-160,477455	331,647545	88,980845	
Octubre	-59,283964	287,674336	460,132736	-59,992264	
Novembre	302,245684	-99,004316	25,203984	-66,754316	
Desembre	-224,748979	-130,415679	209,501021	65,251021	

Per a facilitar la interpretació de les dades d'aquesta taula, hem ressaltat en roig els valors que corresponen a períodes on la utilització del transport públic ha estat per damunt del que s'esperava, atenent al model que hem trobat, i els valors ressenyats en blau corresponen a períodes amb valors que estan sota les previsions, sempre segons el model que hem estudiat.

Podem destacar que els valors que es troben d'alguna manera «fora del model» corresponen, en tots els anys de la nostra taula, als mesos de març i abril. Podríem explicar que, potser, cal considerar que en aquests mesos es produeixen festes importants en la Comunitat i que les vacances de Setmana Santa i Pasqua són també en aquests mesos de manera oscil·lant i no en un mes concret, cosa que haguera facilitat l'estudi de la component estacional.

Fem aquesta consideració perquè hem observat que la component estacional ens permet interpretar que els mesos que corresponen a vacances escolars i/o laborals (juliol, agost, setembre i desembre) tenen valors que indiquen menys utilització del transport públic. També el mes d'abril està en aquest grup i pot variar, segons l'any, el comportament del mes de març i abril com es pot interpretar de les dades de la component erràtica o residual.

Exercici 6

Realitzeu la gràfica de la sèrie següent que indica els milers de quilos de fruita comercialitzada per trimestres al llarg dels darrers 4 anys.

	2008	2009	2010	2011
1r trimestre	10	23	12	29
2n trimestre	11	27	11	28
3r trimestre	9	25	10	21
4t trimestre	8	20	8	23

Observeu les taules següents que presenten els càlculs que hem fet per a obtenir la tendència i la component estacional de la sèrie.

dades		
10		
11		
9		
8		
23		
27		
	16,875	
25		16,9375
	17	
20		17
	17	
12		17,375
11		
	17,875	

dades		
10		
11		
9		
	12,75	
8		
	16,75	
23		18,75
27		
	23,75	
25		22,375
	21	
20		19
	17	
12		15,125
	13,25	
11		11,75
	10,25	

10		17,625
	17,375	
8		
29		
28		
21		
23		

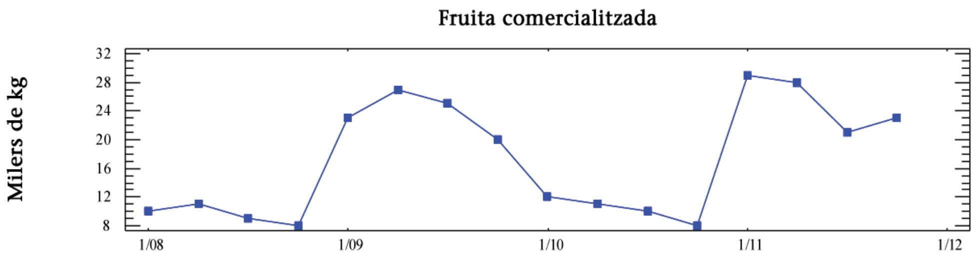
10		
8		16,625
	18,75	
29		20,125
28		
21		
23		

Solució

a) Identifiqueu el mètode emprat, afegiu aquelles dades que falten en les taules, comentant el procediment de càlcul que cal fer-hi, i justifica'l.

Gràfic i visualització del model

Per començar, presentarem el gràfic de les dades de la sèrie original:



i podem veure que la periodicitat del fenomen fa que les dades repetisquen el «patró» cada 2 anys.

Càlcul de la tendència

Ens demanen que identifiquem el procediment per a obtenir les components de la sèrie. Sembla prou evident que estem treballant pel mètode de les mitjanes mòbils i caldrà veure quin és el nombre de dades que cal agafar per a fer cada mitjana en el càlcul de la tendència.

Per a decidir-nos cal considerar dos nombres: el nombre de dades de cada «període» de la gràfica abans presentada (en aquest cas, és evident que és 8) i el nombre d'observacions per any (en aquest cas, és 4 perquè son dades trimestrals).

Per la qual cosa el nombre de dades a agafar és MCM(4, 8) = 8. Però aquesta decisió permet veure que la taula de l'esquerra està construïda per a calcular la tendència, fent en la columna primera les mitjanes amb p = 8, i a la dreta el posterior centrat que cal fer per tractar-se d'un nombre parell de dades.

També observem que hi és incompleta. Presentem la taula completada i a continuació els càlculs que hem fet per a completar-la.

La primera mitjana que falta en la primera columna, l'hem obtingut agafant les primeres 8 dades (ho indicarem al subíndex). Així:

$$y_{1-8} = \frac{10+11+9+8+23+27+25+20}{8} = 16{,}625$$

de la mateixa manera, continuarem omplint les caselles següents de la mateixa columna. Indiquem al subíndex l'ordinal de les dades extremes de l'interval de dades que agafarem en cada mitjana. Cal recordar que estem emprant un mètode de «suavitzat» per mitjanes mòbils.

$$y_{2-9} = \frac{11+9+8+23+27+25+20+12}{8} = 16{,}875$$

$$y_{3-10} = \frac{9+8+23+27+25+20+12+11}{8} = 16{,}875 \text{ (que ja figura a la taula)}$$

i així podem anar comprovant les cel·les ja calculades fins que arribem a la mitjana següent que està buida:

$$y_{6-13} = \frac{27+25+20+12+11+10+8+29}{8} = 17{,}75$$

i també falta la darrera cel·la que correspon a la mitjana següent (que calculem amb les dades dels darrers dos anys):

$$y_{9-16} = \frac{12+11+10+8+29+28+21+23}{8} = 17{,}75$$

Com podem veure en la taula anterior, aquesta columna no correspon encara als valors de la tendència ja que les mitjanes obtingudes no corresponen als períodes (trimestres) dels quals tenim les dades. Com que es tracta de mitjanes obtingudes amb un nombre parell de dades, cal fer un posterior centrat per a evitar aquest problema. A tal fi, cal fer la mitjana aritmètica de cada dues dades (mitjanes de la columna primera).

Per a completar aquesta segona columna, que ja correspon als valors de la tendència per a cada període (T_{ik}), hem omplit les primers caselles que faltaven, a partir de les primeres dades de la columna anterior. Així, tenim:

$$T_{09/1} = \frac{16,625+16,875}{2} = 16,75$$

$$T_{09/2} = \frac{16,875+16,875}{2} = 16,875$$

les dades següents de la taula ja estan calculades fins a arribar a la següent cel·la buida que correspon a:

$$T_{10/2} = \frac{17,75+17,875}{2} = 17,8125$$

i la darrera, obtinguda a partir de les últimes mitjanes:

$$T_{10/4} = \frac{17,375+17,75}{2} = 17,5625$$

Podem veure que el càlcul de la tendència per aquest mètode impedeix obtenir la tendència dels primers i darrers períodes per manca de dades per a calcular les mitjanes mòbils i posterior centrat.

Calcula la component estacional

El càlcul de la component estacional pel mètode de les mitjanes mòbils està reflectit a la taula de la dreta. En aquest cas, el nombre d'observacions que cal agafar per a cada mitjana, és el nombre d'observacions que tenim per any. És a dir, les dades són trimestrals, per tant, cal agafar 4 dades en cada mitjana.

Mostrarem a continuació els càlculs de les mitjanes que falten en les cel·les buides i indicarem al subíndex l'ordinal de les dades de l'interval de valors que agafarem en cadascuna.

Així, començarem per les primeres cel·les que podem calcular en la primera columna. Els resultats estan en roig en la taula següent:

$$\overline{y}_{1-4} = \frac{10+11+9+8}{4} = 9,5$$

$$\overline{y}_{2-5} = \frac{11+9+8+23}{4} = 12,75 \text{ (ja està en la taula)}$$

Podem comprovar els resultats següents fins a arribar a les cel·les:

$$\overline{y}_{4-7} = \frac{8+23+27+25}{4} = 20,75$$

$$\overline{y}_{10-13} = \frac{11+10+8+29}{4} = 14,5$$

i les darreres de la columna:

$$\bar{y}_{12-15} = \frac{8+29+28+21}{4} = 21{,}5$$

$$\bar{y}_{13-16} = \frac{29+28+21+23}{4} = 25{,}25$$

Com que també es tracta d'un nombre parell de dades (p = 4) caldrà també fer el centrat, igual que amb la taula de l'esquerra. Per a omplir la segona columna recordem que s'haurà de fer la mitjana aritmètica de cada dues mitjanes. Indicarem al subíndex l'any i trimestre al qual fer correspondre cada resultat. Així, la primera cel·la, denotada per $\bar{y}'_{08/3}$ correspon a la mitjana mòbil del 3r trimestre de l'any 2008.

$$\bar{y}'_{08/3} = \frac{9{,}5+12{,}75}{2} = 11{,}125$$

$$\bar{y}'_{08/4} = \frac{12{,}75+16{,}75}{2} = 14{,}75$$

$$\bar{y}'_{09/2} = \frac{20{,}75+23{,}75}{2} = 22{,}25$$

$$\bar{y}'_{10/3} = \frac{10{,}25+14{,}5}{2} = 12{,}375$$

$$\bar{y}'_{11/2} = \frac{21{,}5+25{,}25}{2} = 23{,}375$$

Tots aquests resultats es poden veure situats en la taula següent:

dades		T_k
10		
11		
9		
8		
	16,625	
23		**16,75**
	16,875	
27		**16,875**

dades		$\bar{y}'_k$
10		
11		
	9,5	
9		**11,125**
	12,75	
8		**14,75**
	16,75	
23		18,75
	20,75	
27		**22,25**

	16,875	
25		16,9375
	17	
20		17
	17	
12		17,375
	17,75	
11		**17,8125**
	17,875	
10		17,625
	17,375	
8		**17,5625**
	17,75	
29		
28		
21		
23		

	23,75	
25		22,375
	21	
20		19
	17	
12		15,125
	13,25	
11		11,75
	10,25	
10		**12,375**
	14,5	
8		16,625
	18,75	
29		20,125
	21,5	
28		**23,375**
	25,25	
21		
23		

Però, recordem que aquestes mitjanes de la columna dreta de la taula dreta no són més que un dels passos del càlcul de la component estacional:

Cal fer, a banda, el càlcul de les mitjanes de les dades originals per períodes. Les denotarem per $\bar{y}_k$ i es troben a la columna dreta de la taula següent:

	2008	2009	2010	2011	$\bar{y}_k$
1r trimestre	10	23	12	29	18,5
2n trimestre	11	27	11	28	19,25
3r trimestre	9	25	10	21	16,25
4t trimestre	8	20	8	23	14,75

Per altra banda, situarem les mitjanes mòbils de la taula de la dreta, ja complimentada, en aquesta distribució per anys i períodes, fent correspondre cada valor al període corresponent. Així:

	2008	2009	2010	2011	$\overline{y}_k$	$\overline{E}_k$	e_k
1r trimestre		18,75	15,125	20,125	18,5	18	0,5
2n trimestre		22,25	11,75	23,375	19,25	19,125	0,125
3r trimestre	11,125	22,375	12,375		16,25	15,292	0,958
4t trimestre	14,75	19	16,625		14,75	16,792	-2,042

Mantenim la columna de la taula anterior amb els valors de $\overline{y}_k$ i a la columna següent, també calculem la mitjana aritmètica d'aquests valors per files o períodes, considerant el nombre de valors de què disposem. Cal veure que hi queden cel·les buides. Les denotarem per $\overline{E}_k$.

Anotem a continuació com calcular-les:

$$\overline{E}_1 = \frac{18{,}75 + 15{,}125 + 20{,}125}{3} = 18$$
$$\overline{E}_2 = \frac{22{,}25 + 11{,}75 + 23{,}375}{3} = 19{,}125$$
$$\overline{E}_3 = \frac{11{,}125 + 22{,}375 + 12{,}375}{3} = 15{,}292$$
$$\overline{E}_4 = \frac{14{,}75 + 19 + 16{,}625}{3} = 16{,}792$$

i en la darrera columna ja podem obtenir la component estacional e_k, restant aquestes columnes que acabem d'explicar: $e_k = \overline{y}_k - \overline{E}_k$.

$$e_1 = 18{,}5 - 18 = 0{,}5$$
$$e_2 = 19{,}25 - 19{,}125 = 0{,}125$$
$$e_3 = 16{,}25 - 15{,}292 = 0{,}958$$
$$e_4 = 14{,}75 - 16{,}792 = -2{,}042$$

Podem interpretar, si analitzem aquests resultats, que les dades del 4t trimestre són molt inferiors als de la resta i destacaríem els valors del tercer trimestre per damunt de la mitjana global.

Calcula la component erràtica o residual

Per a calcular la component erràtica o residual, considerant que treballem amb un model additiu, restarem a cada dada original els valors de la tendència i la component estacional. Així, ens distribuirem també la taula dels valors de la tendència per períodes i anys. A la taula següent tenim aquesta distribució, per la qual cosa es pot veure que tan sols podrem calcular-hi la component residual de les dues columnes centrals.

	2008	2009	2010	2011
1r trimestre		16,75	17,375	
2n trimestre		16,875	17,8125	
3r trimestre		16,9375	17,625	
4t trimestre		17	17,5625	

Així, considerant que $r_{ik} = y_k - T_k - e_k$, tenim els resultats següents:

	2008	2009	2010	2011
1r trimestre		23 – 16,75 – 0,5 = 5,75	12 – 17,375 – 0,5 = –5,875	
2n trimestre		27 – 16,875 – 0,125 = 10	11 – 17,8125 – 0,125 = –6,9375	
3r trimestre		25 – 16,9375 – 0,958 = 7,1045	10 – 17,625 – 0,958 = –8,583	
4t trimestre		20 – 17 + 2,042 = 5,042	8 – 17,5625 + 2,042 = –7,5205	

En els valors que obtenim en aquesta taula, hi destaca molt el comportament de les dades amb fluctuacions molts importants de cada any, com podíem veure en la gràfica inicial. És a dir, que un any té els valors molt baixos i el següent molt més grans. Aquesta circumstància és la que podem veure reflectida en la taula anterior de la component residual, que en aquesta sèrie, fa que ressalte el comportament cíclic bianual.

Exercici 7

La sèrie cronològica següent mostra el nombre de noves contractacions de certa superfície comercial per quadrimestres en els anys que s'indiquen:

	2007	2008	2009	2010	2011
1r quadrimestre	41	39	35	21	22
2n quadrimestre	37	33	27	15	16
3r quadrimestre	36	30	16	12	13

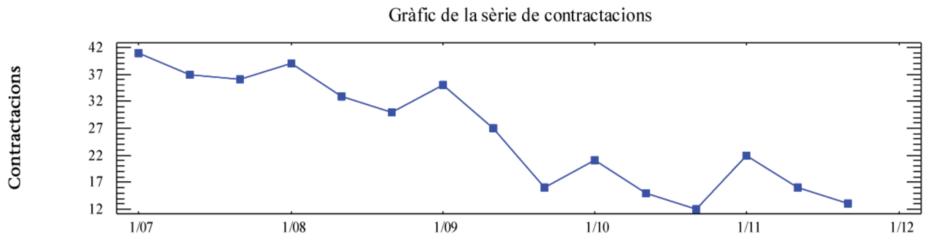
a) Realitzeu la gràfica de la sèrie i suposant un model additiu, calculeu pel mètode de l'ajust analític, les components d'aquesta sèrie i interpreteu cadascun dels resultats obtinguts.
b) Feu les previsions que podem esperar per els anys 2012 i 2013.
c) Realitzeu l'anàlisi de la sèrie pel mètode de les mitjanes mòbils, calculant també les components de la sèrie.
d) Compareu els resultats de l'anàlisi pels dos mètodes emprats.

Solució

a) Realitzeu la gràfica de la sèrie i suposant un model additiu, calculeu pel mètode de l'ajust analític, les components d'aquesta sèrie i interpreteu cadascun dels resultats obtinguts.

Identificació del model i gràfica

Sempre cal començar amb una gràfica de les dades per veure el patró de comportament de la sèrie i que ens confirme que podem aplicar-li un ajust analític de tipus additiu:



En veure aquesta gràfica cal esperar una tendència decreixent i també una component estacional prou marcada per la «periodicitat» que podem advertir en la gràfica (eix X) que coincideix amb l'interval anual, que correspon cada tres dades.

Càlcul de la tendència

Per a calcular la component de la tendència omplirem les caselles de la part inferior de la taula següent per a ajustar una recta de regressió a les mitjanes quadrimestrals anuals que hi figuren en la fila $\overline{y}_i = \dfrac{\sum_k y_{ik}}{m}$.

Cal notar que $\overline{y}_i$ és la mitjana dels valors de cada columna. Recordem tan sols que cada dada d'una sèrie temporal la representem per dos subíndex y_{ik}, on i fa referència a l'any i k fa referència al període. A més, m indica el nombre de files de la taula o nombre de períodes dels quals tenim dades per cada any. En el nostre cas $m = 3$ perquè un any té 3 quadrimestres.

Podem veure-hi que a la fila següent hi ha una escala i que fa referència als anys per a facilitar els càlculs; és recomanable posar el valor 0 en la columna central perquè simplificarà molt el càlcul del sistema que cal plantejar per a obtenir els coeficients de la recta de tendència.

A continuació, omplirem les files tercera i quarta que fan referència a $\overline{y}_i \cdot i$ i i^2, considerant els valors de les files anteriors.

Després sumarem les files i obtindrem els valors que podem trobar en la columna «totals».

	2007	2008	2009	2010	2011	
1r quadrimestre	41	39	35	21	22	
2n quadrimestre	37	33	27	15	16	
3r quadrimestre	36	30	16	12	13	TOTALS
$\overline{y}_i$	38	34	26	16	17	131
i	-2	-1	0	1	2	0
$\overline{y}_i \cdot i$	-76	-34	0	16	34	-60
i^2	4	1	0	1	4	10

Utilitzarem aquests valors «totals» per a resoldre el sistema que es planteja a continuació:

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases}$$

on N és el nombre d'anys dels quals tenim dades en la taula (en aquest cas, 4 anys que corresponen a les 4 columnes), i els coeficients a i b són els coeficients de la recta de regressió que denominem recta de tendència, la fórmula de la qual és $T_i = a + b$, on i fa referència a l'any que s'indica en l'escala de la taula superior.

$$\begin{cases} \sum_i \overline{y}_i = Na + b\sum_i i \\ \sum_i \overline{y}_i \cdot i = a\sum_i i + b\sum_i i^2 \end{cases} \Rightarrow \begin{cases} 131 = 5a + 0b \\ -60 = 0a + 10b \end{cases}$$

Resoldrem el sistema aïllant en cada equació el valor dels coeficients a i b. La tasca es facilita per la situació del 0 de l'escala en la columna central quan el seu nombre és senar.

Els valors que hem calculat són: $a = \frac{131}{5} = 26{,}2$ i $b = \frac{-60}{10} = -6$, amb els quals podem concloure que $T_i = 26,-6i$.

Per a interpretar aquesta component de la sèrie, T_i, que s'anomena tendència i que ens explica el comportament del fenomen a «llarg termini», ens fixem en el valor de la pendent $b = -6$, que per ser negativa ens permet explicar que les contractacions decreixen any rere any, i podem a més a més detallar que el valor de la mitjana de despeses quadrimestrals del departament, $\overline{y}_i$, ha disminuït en 6 el número de contractacions cada any.

Càlcul de la component estacional

La segona component de la sèrie que cal avaluar és la component estacional, que denotarem per e_k, cosa que ens permetrà analitzar el comportament del fenomen per períodes dintre de l'any (en el nostre exercici, per quadrimestres), valorant en quin d'ells els valors estan per damunt o per sota d'un valor global que denotarem per M i que anomenarem mitjana global corregida.

Els valors de e_k estaran expressats en valors absoluts i en les mateixes unitats que les dades de la taula original (en aquest exercici en milers d'euros).

Per a abordar els càlculs de e_k, treballarem ampliant la taula en diferents columnes cap a la dreta de la taula original, ja que ens interessa fer un treball per períodes, en aquest cas, per quadrimestres.

En la columna primera calcularem la mitjana dels valors de cada quadrimestre i aquesta mitjana la denotarem per $\overline{y}_k = \frac{\sum_i y_{ij}}{N}$.

	2007	2008	2009	2010	2011		$\overline{y}_k$	$\overline{y}_k^{'}$	e_k	I_k
1r quadrimestre	41	39	35	21	22		31,6	31,6	3,4	112,06%
2n quadrimestre	37	33	27	15	16		25,6	27,6	–0,6	97,87%
3r quadrimestre	36	30	16	12	13	TOTALS ↓	21,4	25,4	–2,8	90,07%
								M = 28,2		

En la columna següent calcularem les mitjanes corregides $\overline{y}_k^{'}$ (on eliminem en cada dada el valor proporcional de la tendència que li podem assignar a cada període), suposant que aquest decreixement b = –6 ha estat constant al llarg dels períodes de l'any.

En aquest exercici, $b/m = -6/3 = -2$. Com que és una tendència negativa, els valors de les mitjanes corregides $\overline{y}_k^{'}$ seran majors que les mitjanes originals $\overline{y}_k$, ja que s'hi calculen així:

$$\overline{y}_k^{'} = \overline{y}_k - \frac{b}{m}(k-1)$$

D'aquesta manera:

$$\overline{y}_1' = \overline{y}_1 - \frac{b}{m}(1-1) = 31,6 - (-2)(1-1) = 31,6$$

$$\overline{y}_2' = \overline{y}_2 - \frac{b}{m}(2-1) = 25,6 - (-2)(2-1) = 27,6$$

$$\overline{y}_3' = \overline{y}_3 - \frac{b}{m}(3-1) = 21,4 - (-2)(3-1) = 25,4$$

A continuació, en una cel·la inferior, calculem *M*, la mitjana de les mitjanes corregides, ja que:

$$M = \frac{\sum_k \overline{y}_k'}{m} = \frac{31,6 + 27,6 + 25,4}{3} = 28,2$$

i representa el valor de referència per a analitzar els valors dels períodes, mitjançant $e_k = \overline{y}_k' - M$, que calcularem en la següent columna.

$$e_1 = \overline{y}_1' - M = 31,6 - 28,2 = 3,4$$

$$e_2 = \overline{y}_2' - M = 27,6 - 28,2 = -0,6$$

$$e_3 = \overline{y}_3' - M = 25,4 - 28,2 = -2,8$$

La component estacional e_k ens explicita el comportament per períodes en valors absoluts, indicant-nos signe i quantitat en les mateixes unitats que les dades originals (milers d'euros). Així, direm que en el primer quadrimestre les contractacions de la nostra sèrie tenen uns valors per damunt de la mitjana, mentre que les contractacions del segon quadrimestre són lleugerament inferiors. El valor de la mitjana de referència seria M = 28,2 que podria representar una mitjana de contractacions quadrimestrals global.

Una altra interpretació d'aquestes dades es pot donar amb els índexs estacionals $I_k = \frac{\overline{y}_k}{M} \cdot 100$ que calculem en la columna següent on s'indica en forma de percentatge (valor relatiu, on *M* representa el 100%) la mateixa informació que la component estacional, però com que té caràcter percentual es més fàcil de presentar sense particularitzar i donar el valor de *M*.

$$I_1 = \frac{\overline{y}_1'}{M} \cdot 100 = \frac{31,6}{28,2} \cdot 100 = 112,05\%$$

$$I_2 = \frac{\overline{y}_2'}{M} \cdot 100 = \frac{27,6}{28,2} \cdot 100 = 97,87\%$$

$$I_3 = \frac{\overline{y}_3'}{M} \cdot 100 = \frac{25,4}{28,2} \cdot 100 = 90,07\%$$

Aquesta component estacional expressada en termes de percentatge, els índexs estacionals, ens permet analitzar i interpretar el comportament de les dades de la sèrie, és a dir, els valors de les contractacions per quadrimestres.

Podrem afirmar que les despeses eren majors en el primer quadrimestre, amb valors d'un 12,05% superiors a la mitjana anual, mentre que els valors del tercer quadrimestre tan sols arriben a tenir-hi valors inferiors en un 10% aproximadament. Cal destacar que les contractacions del segon quadrimestre tenen valors que estan al voltant de la mitjana anual global, que podríem considerar el valor de $M = 28{,}2$ contractacions.

Càlcul de la component residual o erràtica

En tercer lloc, s'ha de calcular la component erràtica o residual, que denotem per r_{ik}, que ens permet destacar el comportament d'alguna dada y_{ik}, el valor de la qual no es puga explicar per les anteriors components, la qual cosa permetrà inferir que per qualsevol causa a identificar (motius extraordinaris) aquest valor no està dintre del patró de comportament que hi hem trobat i amb el qual hem interpretat les dades originals per explicar el fenomen. Els valors d'aquesta component erràtica cal que siguen menuts en valor, variats en signe i sense cap regularitat ni patró. Ens permeten veure que les dades reals no s'ajusten completament al patró que hem trobat en la tendència i la component estacional. Es per això que si algun valor és molt alt o baix deixarà identificar alguna dada que correspon al període d'un any, el valor de la qual s'allunya molt dels valors que caldria esperar.

La calcularem així $r_{ik} = y_{ik} - T_i - e_k$ en cada cel·la de la taula. Per a tal fi, primerament cal determinar la tendència per a cada any de la taula $T_i = 26{,}2 - 6i$.

$$T_{2007} = T_{-2} = 26{,}2 - 6\cdot(-2) = 38{,}2$$
$$T_{2008} = T_{-1} = 26{,}2 - 6(-1) = 32{,}2$$
$$T_{2009} = T_0 = 26{,}2 - 6\cdot 0 = 26{,}2$$
$$T_{2010} = T_1 = 26{,}2 - 6\cdot 1 = 20{,}2$$
$$T_{2011} = T_2 = 26{,}2 - 6\cdot 2 = 14{,}2$$

Els valors de la component estacional es troben en la columna de la taula e_k, així:

$e_1 = 3{,}4$ $\qquad e_2 = -\,0{,}6$ $\qquad e_3 = -\,2{,}8$

A més, els càlculs per a cada cel·la apareixen en la taula següent:

	2007	2008	2009	2010	2011
1r quadrimestre	41 – 38,2 – 3,4 = –06	39 – 32,2– 3,4 = 3,4	35 – 26,2 – 3,4 = 5,4	21 – 20,2 – 3,4 = –2,6	22 – 14,2 – 3,4 = 4,4
2n quadrimestre	37 – 38,2 + 0,6 = –0,6	33 – 32,2 + 0,6 = 1,4	27– 26,2 + 0,6 = 1,4	15 – 20,2 + 0,6 = –4,6	16 – 14,2 + 0,6 = 2,4
3r quadrimestre	36 – 38,2 + 2,8 = 0,6	30 – 32,2 + 2,8 = 0,6	16 – 26,2 + 2,8 = –7,4	12 – 20,2 + 2,8 = –5,4	13 – 14,2 + 2,8 = 1,6

D'aquests resultats, podem observar que els valors que potser ens criden més l'atenció els remarquem amb color roig. Podrem destacar els corresponents al primer quadrimestre del 2009 que és superior al que caldria esperar amb $r_{2009,1} = 5{,}4$ i el tercer quadrimestre del mateix any amb un valor encara molt inferior amb $r_{2011,3} = -7{,}4$. I també el valor del tercer quadrimestre del 2010 $r_{2010,3} = -5{,}4$ que ens indica que és un valor inferior al que caldria esperar. S'hauria d'estudiar si alguna circumstància extraordinària justifica aquests valors.

b) Estima els valors de les despeses que es poden esperar per a l'any 2012 i 2013.

Per a fer les estimacions que ens demanen en aquest apartat, recordem que no podem preveure la component residual, per la qual cosa calcularem:

$$y_{ik} = T_i + e_k$$

Considerant l'escala que hem adoptat en la taula per als anys, amb el propòsit de simplificar els càlculs de la tendència, podem establir que si $2009 \Rightarrow i = 0$, això comporta $2012 \Rightarrow i = 3$, la qual cosa ens permetrà obtenir el valor de la tendència per a aquest any: $T_{2012} = T_3 = 26{,}2 - 6{\cdot}3 = 8{,}2$ i com que coneixem la component estacional: $e_1 = 3{,}4$, $e_2 = -0{,}6$ i $e_3 = -2{,}8$, podrem estimar els valors de les despeses de l'any 2012, per quadrimestres.

$y_{2012,1qua} = 8{,}2 + 3{,}4 =$ 1,6 contractacions

$y_{2012,2qua} =$ 8,2–0,6 = 7,6 contractacions

$y_{2012,3qua} =$ 8,2–2,8 = 5,4 contractacions

Repetirem el procés per a l'any $2013 \Rightarrow i = 4$ amb el valor de la tendència $T_{2013} = T_4 = 26{,}2 - 6{\cdot}4 = 2{,}2$ i podrem estimar:

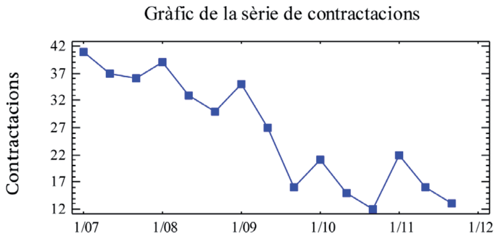
$y_{2013,1qua} = 2{,}2 + 3{,}4 = 5{,}6$ contractacions

$y_{2013,2qua} = 2{,}2 - 0{,}6 = 1{,}6$ contractacions

$y_{2013,3qua} = 2{,}2 - 2{,}8 = -0{,}6$ contractacions *c*) Realitzeu l'anàlisi de la sèrie pel mètode de les mitjanes mòbils, calculant també les components de la sèrie.

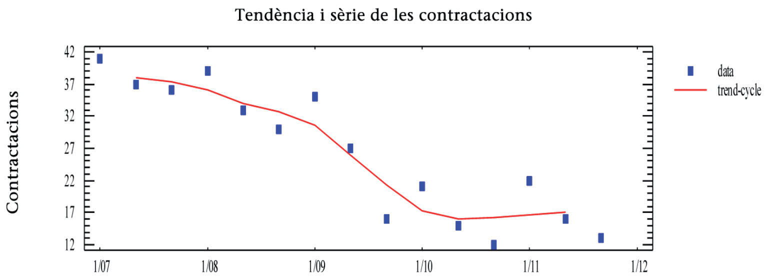
Gràfic i anàlisi de la sèrie

Per a analitzar la sèrie pel mètode de les mitjanes mòbils, cal fixar-se en la gràfica de la sèrie que ens permet veure que té una periodicitat anual.



Càlcul de la tendència

Per a calcular la tendència, cal triar el nombre de dades que agafarem per a determinar cadascuna de les mitjanes. En aquest cas, $p = 3$. A tal fi escrivim en la taula següent els resultats de les mitjanes mòbils que són els valors de la tendència i la gràfica de la seua interpretació:



dades	T_k
41	
37	38,00
36	37,33
39	36,00
33	34,00
30	32,67
35	30,67
27	26,00
16	21,33
21	17,33
15	16,00
12	16,33
22	16,67
16	17,00
13	

S'hi veu un comportament a llarg termini decreixent, encara que amb les darreres dades sembla que comença a corregir-se.

Càlcul de la component estacional

Per a calcular la component estacional, s'han de determinar en primer lloc les mitjanes de les dades originals per períodes. Les denotem per $\bar{y}_k$.

	2007	2008	2009	2010	2011	$\bar{y}_k$
1r quadrimestre	41	39	35	21	22	31,6
2n quadrimestre	37	33	27	15	16	25,6
3r quadrimestre	36	30	16	12	13	21,4

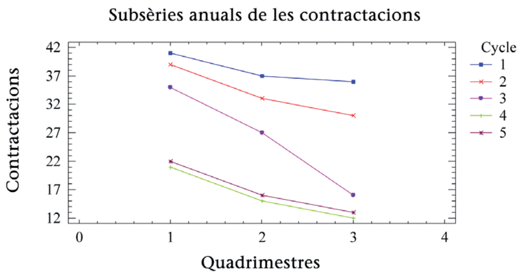
Per altra banda, agafarem les mitjanes mòbils amb $p = 3$, ja que és el nombre de dades per any. En el nostre cas podrem aprofitar les dades ja calculades en l'apartat anterior. Les hi col·locarem distribuïdes per anys i quadrimestres.

En la columna de la dreta afegim les mitjanes de cada quatrimestre per files i la denotem per $\bar{E}_k$.

	2007	2008	2009	2010	2011	$\bar{E}_k$	e_k
1r quadrimestre		36,00	30,67	17,33	16,67	25,17	6,43
2n quadrimestre	38,00	34,00	26,00	16,00	16,56	26,11	–0,51
3r quadrimestre	37,33	32,67	21,33	16,33		26,92	–5,52

Per a calcular la component estacional, restem aquestes columnes calculades entre si. Així, $e_k = \bar{y}_k - \bar{E}_k$ ens permet interpretar el comportament de la sèrie per quadrimestres.

És evident que els primers quadrimestres els valors de la sèrie tenen els majors valors amb 6,43 contractacions per damunt d'una hipotètica mitjana anual. Destacarem també el resultat del tercer quadrimestre amb valors inferiors, que es reflecteixen amb el –5,52 contractacions que veiem en la cel·la corresponent. Aquesta interpretació també és coherent amb la gràfica de la sèrie que hem vist en començar el problema.



Càlcul de la component erràtica o residual

Per a calcular la component residual, restarem a les dades originals la tendència i la component estacional, considerant que no ho podrem fer amb la primera i darrera dada per manca de les mitjanes de la tendència: $r_{ik} = y_{ik} - T_{ik} - e_k$. En la taula següent presentem els càlculs i resultats.

	2007	2008	2009	2010	2011
1r Quad.		39 – 36 – 6,43 = –3,43	35 – 30,67 – 6,43 = –2,1	21 – 17,33 – 6,43 = –2,76	22 – 16,67 – 6,43 = –1,1
2n Quad.	37 – 38 + 0,51 = –0,49	33 – 34 + 0,51 = –0,49	27 – 26 + 0,51 = 1,51	15 – 16 + 0,51 = –0,49	16 – 16,56 + 0,51 = –0,05
3r Quad.	36 – 37,33 + 5,52 = 4,19	30 – 32,67 + 5,52 = 2,85	16 – 21,33 + 5,52 = 0,19	12 – 16,33 + 5,52 = 1,19	

Encara que no són quantitats gaire elevades, hem ressenyat en roig les dades que corresponen al 1r quadrimestre de l'any 2008, per ser 3,43 menys contractacions del que s'espera per la regularitat del fenomen. També hem detallat la dada del 3r quadrimestre del 2007 que és 4,19 contractacions per damunt.

d) Compareu els resultats de l'anàlisi pels dos mètodes emprats.

La principal diferència entre ambdós mètodes és la possibilitat de fer previsions pel mètode de l'ajust, que no podem fer en el de les mitjanes mòbils, ja que l'equació de la recta de tendència que obtenim al mètode primer enunciat, ens permet obtenir valors de tendència per anys propers, sempre que considerem com a hipòtesi de treball que el fenomen mantindrà un comportament semblant a les dades reals que tenim.

En ambdós mètodes la tendència és decreixent.

Si considerem la component estacional del mètode de l'ajust

$e_1 = 3{,}4 \qquad e_2 = -\,0{,}6 \qquad e_3 = -\,2{,}8$

i les comparem amb les del mètode de les mitjanes mòbils

$e_1 = 6{,}43 \qquad e_2 = -0{,}51 \qquad e_3 = -5{,}52$

podem observar que no coincideixen en valors absoluts però sí en l'apreciació subjectiva i els valors relatius donen la mateixa interpretació.

En la component residual es poden veure resultats molt diferents. També cal veure que les dades dels darrers dos anys són molt inferiors als anteriors i poden «trencar» un poc el model. Ara bé, com el comportament estacional sí era constant, aquesta component ha estat més clara.

Bibliografia

Barbancho, A. G., *Estadística Elemental Moderna.* Ed. Ariel Economía, 1982.

Beltrán, M. J. y Peris, M. J., *Introducció a l'estadística aplicada a les ciències socials* Ed. Sapientia del servei de publicacions de la uji,, número 24.

Escudero-Valles, R., *Métodos Estadísticos Aplicados a la Economía.* Ed. Ariel Economía, 1994.

Biosca, A.; Espinet, M. J.; Fandos, M. J.; Jimeno, M.; Villagrà, J., *Matemáticas aplicadas a las Ciencias Sociales II*. Barcelona: Edebé, 1999.

Brunet, I., Belzunegui, A. y Pastor, I. *Les tècniques d'investigació social i la seva aplicació.* Universitat Rovira i Virgili, 2000.

Colera, J.; García, R.; Oliveira, M. J. *Matemàtiques aplicades a les Ciències Socials*. Madrid: Anaya, 2003.

Correa, J. C. y González, N., *Gráficos en R*. Universidad Nacional Sede Medellín, 2002.

Fernández Cuesta, C. y Fuentes García, F., *Curso de Estadística Descriptiva. Teoria y pràctica*. Ed. Ariel 1994.

Gracia, F.; Mateu, J.; Vindel, P., *Problemas de Probabilidad y Estadística.* Valencia: Tilde, 1997.

Ibáñez, M. V.; Simó, A., *Apuntes de Estadística para Ciencias Empresariales*. Castelló: uji, 2002.

Kazmier, L., *Estadística aplicada a la administración y a la economía*. Ed. MC Graw Hill, 3.ª ed., 1998.

Martín, P. y Martín Pliego, J., *Curso Básico de Estadística Económica*. Ed. AC (3.ª ed.) 1991.

Martín Pliego, J., *Introducción a la Estadística Económica y Empresarial*. Ed. AC. Colección Plan Nuevo, 2004.

Meyer, P. L. *Probabilidad y aplicaciones estadísticas,* Ed. Addson-Wesley, 1986.

Monteagudo, M. F. y Paz, J., *Matemáticas aplicadas a las Ciencias Sociales II*. Zaragoza: Luis Vives, 2003.

Montero Lorenzo, J. M., *Estadística para Relaciones Laborales*. Ed. AC, 2003.

Newbold, P. Carlson, W. L. y Thorne, B., *Estadística para administración y economía*. Ed. Prentice Hill, 2007.

Ruíz-Maya Pérez, L., Martín-Pliego López, F. J., *Fundamentos de inferencia Estadística,* 3.ª ed., Thomson, 2005.

Sanz, J. A., Bedate, A., Rivas, A. y González, J., *Problemas de Estadística descriptiva empresarial.* Ed. Ariel Economía, 1996.

Spiegel, M., *Estadística.* Ed. Mc. Graw-Hill. Sèrie Schaum, 1970.

Tomeo Perucha, V. y Uña Juárez, I., *Diez Lecciones de Estadística Descriptiva (Curso Teórico-Práctico).* Ed. AC, 2003.

Triola, M. F. *Estadística Elemental*, Ed. Pearson Educations, 7.ª ed., 2000.

Venables, W. N., Smith , D. M. y The R Development Core Team, *An introduction to R*. ISBN 3-900051-12-7, 2008.

Webster, A. L., *Estadística aplicada a los negocios y a la economia*. Ed. MC Graw-Hill, 2000.

Wonnacot, T. H. y Wonnacot, R. J., *Introducción a la Estadística*. Limusa Noriega Editores, 1996.

Zaiats, V., Calle, M. L. y Presas, R., *Probabilitat i Estadística. Exercicis* I. Ed. Eumo, 1998.

Webs

www.ine.es
www. ub.edu/stat/GrupsInnovació/Statmedia
www.bioestadistica.uma.es/libro
www.monografias.com