

Examen 6/2/2004. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado. MODELO A.

1. (1 punto) Se han probado dos algoritmos de reenvío de paquetes. Se ha comprobado que el algoritmo  $A$  perdía 5 de 1000 paquetes, mientras que el  $B$  6 de 1000.
  - a) Encuentra un intervalo de confianza al 95 % para la diferencia entre las proporciones de pérdidas de ambos algoritmos e interprétalo.
  - b) Consideremos ahora únicamente el algoritmo  $A$ . Queremos estimar la proporción de paquetes perdidos de este algoritmo, y deseamos estar al menos 95 % seguros que el error es como mucho de 0.001. ¿Cómo ha de ser de grande la muestra si
    - (i) usamos la proporción anterior como una muestra preliminar que nos proporciona una primera estimación?
    - (ii) no tenemos idea de cuál pueda ser la proporción real?
2. (1.75 puntos) Se quiere comparar la rapidez de dos modelos de impresora  $A$  y  $B$ . Los de la compañía  $A$  sostienen que el tiempo medio de impresión de su modelo es más de 5 segundos más rápido que para el modelo  $B$  de los rivales. Se mide el tiempo de impresión de los dos modelos (que consideraremos normal) sobre una serie de 8 plantillas estándar y los resultados aparecen en la tabla siguiente:

	Tiempo para $A$	Tiempo para $B$
Plantilla 1	20	26
Plantilla 2	25	29
Plantilla 3	22	27
Plantilla 4	23	28
Plantilla 5	19	29
Plantilla 6	21	30
Plantilla 7	18	25
Plantilla 8	20	26

- a) Utiliza el contraste adecuado para comprobar si este estudio confirma la afirmación de la compañía  $A$  a nivel de significación de 0.05.
  - b) Si ahora se considerara únicamente una muestra aleatoria de los tiempos de impresión para 70 plantillas para la impresora  $A$ , de las que se obtiene una media de 20 y una desviación típica de 2. Calcula el intervalo de confianza al 95 % para el tiempo medio de impresión de la impresora  $A$ .
  - c) ¿Cuál es la interpretación del intervalo anterior?
  - d) ¿Es necesario que la distribución sea Normal para que el intervalo anterior sea válido? (razona la respuesta).
  - e) Si deseamos que el error en la estimación de la media anterior sea inferior a 0.35 segundos con una confianza del 95 %, y teniendo en cuenta que podríamos asumir  $\sigma = 2$ , ¿cuál será el tamaño muestral requerido?
3. (1.5 puntos) Un grupo de investigación en el rendimiento de las comunicaciones por red quiere saber si una implementación de un protocolo de comunicación de redes envía los paquetes con el "payload" (información útil) que se quiere transmitir o si realmente añade alguna información adicional para que los paquetes tengan una medida mínima. Para ello, han hecho unas pruebas donde enviaban paquetes con 3 diferentes longitudes de "payload" y calculaban los tiempos que tardaban desde que los enviaban hasta que los recibían en el servidor. Todas las pruebas se han realizado bajo las mismas condiciones y se han enviado

5 paquetes de cada una de las medidas (100, 200 y 300 bytes). Las observaciones (tiempo en milisegundos) se recogen en la siguiente tabla:

Longitud (bytes)	Observaciones				
100 (Longitud 1)	38	40	36	39	41
200 (Longitud 2)	51	50	50	49	52
300 (Longitud 3)	53	51	50	49	55

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas en el tiempo medio en función de la longitud de los ficheros.

ANOVA Table for Tiempo by Longitud

Analysis of Variance				
Source	Sum of Squares	Df	Mean Square	F-Ratio
Between groups				
Within groups			3,6	
Total (Corr.)	542,933			

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados.
4. (1.25 puntos) Tras observar un sistema informático se ha obtenido la siguiente distribución de la variable  $X = \text{"Número de colisiones al buscar una dirección de memoria libre"}$ :

$X = \text{Colisiones}$	0	1	2	3	4	5	$\geq 6$
Frecuencia observada	229	211	93	35	7	1	0

Decide al 95 % si estos datos siguen una distribución de Poisson.

5. (2 puntos) Se controla el proceso de envasado de cereales para el desayuno, utilizando los gráficos  $\bar{X}$  y  $R$ . Se toman 20 muestras de tamaño 5, obteniéndose los siguientes datos:

$$\sum_{i=1}^{20} \bar{x}_i = 326'4, \text{ y } \sum_{i=1}^{20} r_i = 15'7.$$

- Encuentra los límites de control de prueba de los gráficos  $\bar{X}$  y  $R$ .
- Suponiendo que el proceso está bajo control, estima la media y desviación típica del proceso.  
Supongamos que las cajas tienen que pesar entre 15'7 y 17.
- ¿Cuál es el porcentaje de unidades defectuosas que se está produciendo en este proceso?
- Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.  
Supongamos que la media se corre a 16.
- ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento?
- ¿Cuál es la ARL después del corrimiento?

Examen 11/9/2004. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado.

1. Cierta programación comercial de traducción automática que traduce del inglés al italiano se apoya en un diccionario bilingüe electrónico que llamaremos A. Un equipo de investigadores en procesamiento del lenguaje natural se propone probar la efectividad de este diccionario.

a) (0.75 puntos) Para ello coge un texto de 1545 palabras que se pasa por el bilingüe electrónico. La traducción propuesta con el diccionario se valida con un filólogo, que considera correcta la traducción de 1154 términos.

(i) Encuentra un intervalo de confianza al 95 % para la proporción con la que el diccionario propone una traducción incorrecta.

(ii) ¿Cuál es la interpretación del intervalo anterior?

(iii) Si deseamos que el error en la estimación de la proporción anterior sea inferior a 0.015 con una confianza del 95 %, y usando la proporción anterior como una muestra preliminar que nos proporciona una primera estimación, ¿cuál será el tamaño muestral requerido?

b) (1.75 puntos) El equipo de investigadores mencionado está desarrollando un segundo bilingüe inglés-italiano, que sea más preciso que el anterior. Se han hecho pruebas con dos métodos alternativos B y C, que parecen más precisos que A. Sin embargo, parece que el incremento en la precisión supone un mayor consumo de espacio de memoria. Por ello, quiere verse como se comporta la ocupación de memoria de cada diccionario.

Para tres muestras aleatorias independientes de textos de igual tamaño, se ha hecho la traducción empleando el método A, B y C sobre un PC aislado y se ha recogido el espacio de memoria (en miles de bytes) consumidos en cada ejecución, obteniéndose los siguientes resultados:

Diccionario	Observaciones			
A	1803	1850	1700	1940
B	2105	2010	2180	2120
C	2290	2098	2240	2310

A partir de estos datos se ha construido la siguiente tabla ANOVA, para estudiar si hay diferencias significativas en la ocupación media de memoria para procesar los datos.

ANOVA Table for Memoria by Diccionario

Analysis of Variance				
Source	Sum of Squares	Df	Mean Square	F-Ratio
Between groups				
Within groups			8028.72	
Total (Corr.)	425462			

(i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.

(ii) Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados.

(iii) Si en el experimento hubiésemos procesado la misma muestra de ficheros con los tres diccionarios, ¿qué prueba habrías utilizado?

2. (2 puntos) Un informático ha de implementar un proceso que gestione una cola. Se plantea la posibilidad de utilizar la política FIFO (first in first out) o bien LIFO (last in first out). Para compararlas, diseña un prototipo implementando cada una de las dos políticas y mide los tiempos de respuesta, en milisegundos, para cada algoritmo, en su sistema. Para diversas ejecuciones se ha observado el tiempo de respuesta, que asumiremos normal, obteniéndose:

Cola	Tamaño muestral	Media muestral	Desviación típica muestral
FIFO	6	5.3	0.9
LIFO	8	4.7	1.1

- a) Calcula un intervalo de confianza al 95 % para el cociente de varianzas y determina (razonando porqué) si existe diferencia entre ellas.
- b) ¿Puede concluirse al nivel  $\alpha = 0.01$  que el tiempo de respuesta medio con FIFO es mayor que con LIFO? Utiliza el contraste adecuado para responder la pregunta anterior. (Si no has resuelto el apartado a), asume igualdad de varianzas).
3. (1.5 puntos) Una empresa del Silicon Valley tiene 3 ordenadores en su centro informático que hacen 3 turnos al día. El encargado de mantenimiento de los ordenadores ha hecho una tabla, por ordenador y turno, donde recoge el número de veces que ha sido necesario reiniciar un determinado ordenador durante el turno.

Turno	Ordenador		
	A	B	C
1	10	20	15
2	20	15	20
3	30	10	10

El encargado quiere saber a un nivel de significación del 5 % si el hecho de reiniciar un ordenador es independiente del propio ordenador y del turno.

4. (1.5 puntos) Se controla el grosor de circuitos electrónicos (en pulgadas), que es un importante factor de calidad, utilizando los gráficos  $\bar{X}$  y  $R$ . Se toman 20 muestras de tamaño 5, obteniéndose los siguientes datos:

$$\sum_{i=1}^{20} \bar{x}_i = 1'2, \text{ y } \sum_{i=1}^{20} r_i = 0'27.$$

- a) Encuentra los límites de control de prueba de los gráficos  $\bar{X}$  y  $R$ .
- b) Suponiendo que el proceso está bajo control, estima la media y desviación típica del proceso. Supongamos que el grosor tiene que estar entre 0'055 y 0'076.
- c) ¿Cuál es el porcentaje de unidades defectuosas que se está produciendo en este proceso?
- d) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos. Supongamos que la media se corre a 0'065.
- e) ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento? ¿Cuál es la ARL después del corrimiento?

Examen 26/1/2005. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado. MODELO B.  
 En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

- (1.5 puntos) Una empresa fotográfica tiene que realizar una compra de impresoras de gran calidad que se van a utilizar en imprimir fotografías digitales. La empresa tiene ofertas de 4 marcas de impresoras de similares características y precio. Para la empresa fotográfica es muy importante la "velocidad de impresión" y, por este motivo, está interesada en saber si las 4 impresoras ofertadas tienen la misma velocidad o hay una que es más rápida. Para responder a esta pregunta decide hacer un experimento que consiste en elegir una única muestra de 5 fotos e imprimirlas en las 4 impresoras. Los resultados del experimento se recogen en la tabla siguiente:

Marca de impresora	Fotos				
	1	2	3	4	5
1	93	89	97	94	88
2	85	77	92	79	84
3	80	87	87	85	81
4	86	81	80	88	79

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas en el tiempo medio de impresión en función de la marca de impresora.

Analysis of Variance for Tiempo - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio
MAIN EFFECTS				
A:Marcaimpresora				
B:Foto	94,8			
RESIDUAL			17,0	
TOTAL (CORRECTED)	592,8			

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - Calcula la LSD al 95% y realiza las comparaciones explicando los resultados.
- (1.25 puntos) Estamos interesados en comprobar la perfección de un dado cúbico (un dado normal de 6 caras), es decir, en comprobar si los resultados se distribuyen uniformemente. Con los resultados obtenidos en 60 lanzamientos del dado, decide si se distribuirían uniformemente usando  $\alpha = 0.05$ :

Resultado	1	2	3	4	5	6
Frecuencia	15	9	7	13	12	4

3. (2.75 puntos) En una multinacional que se dedica a la venta de pantallas, se consideran dos modelos. El departamento de ingeniería ha realizado pruebas de duración para los modelos, que se recogen a continuación:

Modelo viejo	9600	8400	8500	9500	9400	8300
Modelo nuevo	10300	9900	10000	9800	10200	

- a) Calcula un intervalo de confianza al 95 % para el cociente de varianzas y determina (razonando porqué) si existe diferencia entre ellas.
  - b) ¿Puede concluirse al nivel  $\alpha = 0.05$  que la duración media del modelo nuevo es 800 horas superior que para el modelo viejo? Utiliza el contraste adecuado para responder la pregunta anterior. (Si no has resuelto el apartado a), asume varianzas distintas).
  - c) Centrándonos ahora únicamente en el modelo nuevo, calcula el intervalo de confianza al 95 % para su duración media.
  - d) Especifica qué elementos podrías modificar, si se desea obtener un intervalo más estrecho en el apartado anterior.
  - e) Señala si es correcta o no la siguiente afirmación, razonando tu respuesta: "  $\mu$  se encuentra en el intervalo obtenido en c), con una probabilidad de 0.95".
4. (2 puntos) Las botellas de plástico para un detergente líquido se hacen por moldeo de soplado.
- a) Se inspeccionan 20 muestras preliminares de 100 botellas cada una, recogándose el número de defectuosas: 12 8 10 7 12 8 9 21 10 5 18 10 12 11 20 3 9 5 9 13.
- i) Utilizando todos los datos, calcula los límites de control para una gráfica P.
  - ii) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- b) Si en el proceso anterior, consideráramos como característica de calidad a controlar, el número de defectos, ¿qué gráfica usarías? ¿por qué?
- c) Consideremos ahora que controlamos el contenido de detergente líquido en las botellas, tomando cada vez muestras de tamaño 7. Supongamos que la media y desviación típica del proceso bajo control fuera respectivamente: 2.05 y 0.3, y los límites del gráfico  $\bar{X}$ : 1.71 y 2.39. Si las especificaciones son  $2 \pm 0.1$ :
- i) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.  
Supongamos que la media se corre a 1.85.
  - ii) ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento? ¿Cuál es la ARL después del corrimiento?

Examen 9/9/2005. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

- (1.5 puntos) Cierta empresa está interesada en conocer la eficiencia de tres programadores. Por ello, analiza el tiempo (en horas) que ha invertido cada programador en resolver 5 warnings diferentes para cada uno, pero similares entre ellos. Los datos aparecen en la siguiente tabla:

Programador	Observaciones				
Programador 1	2.7	3.1	3	2.5	3.6
Programador 2	2.2	2.9	2.8	3.2	2.2
Programador 3	4.5	4.1	4.9	4.7	4.4

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas en el tiempo medio empleado en resolver los warnings según los programadores.

ANOVA Table for Tiempo by Programador

Analysis of Variance				
Source	Sum of Squares	Df	Mean Square	F-Ratio
Between groups				
Within groups			0.155667	
Total (Corr.)	11.7573			

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados.
- (2 puntos) Se toma una muestra de estudiantes universitarios de informática de dos universidades y se les pregunta por su navegador de internet favorito, resultando que en la universidad A, de 200 encuestados 30 prefieren el *Exploder*, mientras que en la universidad B, de 300 encuestados 60 lo prefieren.
    - Encuentra un intervalo de confianza al 95 % para la diferencia entre las proporciones 'de preferencia del *Exploder*' de ambas universidades e interprétalo.
    - Consideremos ahora únicamente la universidad A. Queremos estimar la proporción de 'de preferencia del *Exploder*', y deseamos estar al menos 95 % seguros que el error es como mucho de 0.03. ¿Cómo ha de ser de grande la muestra si usamos la proporción anterior como una muestra preliminar que nos proporciona una primera estimación?
    - Centrándonos de nuevo en la universidad A, los 200 encuestados pertenecían a tres carreras distintas: ITIG, ITIS e II. En la encuesta tenían dos opciones de navegador a elegir: *Exploder* y *Notescapes*, de forma que los datos desglosados aparecen en la tabla siguiente:

Navegador	Carrera		
	ITIG	ITIS	II
Exploder	30	25	15
Notescapes	50	55	25

Nos interesa saber a un nivel de significación del 5 % si la preferencia por un navegador u otro es independiente de la carrera.

3. (2 puntos) La fiabilidad del software es un atributo de calidad extensamente estudiado. Podría definirse como la probabilidad que un sistema ejecute satisfactoriamente las funciones requeridas en un período de tiempo especificado. Entre las métricas usadas para describirla destaca el tiempo de reparación del software. Una empresa de software, para cierto producto, se compromete a realizar telemantenimiento. Un cliente quiere la garantía que, en general, este tiempo no sobrepase los 100 minutos. Para poder asumir normalidad, trabajaremos, sin embargo, con el logaritmo de los datos, de forma que tenemos las siguientes observaciones del logaritmo del tiempo de reparación:  $\log_{10}(79) = 1.898$ ,  $\log_{10}(32) = 1.505$ ,  $\log_{10}(50) = 1.699$ ,  $\log_{10}(40) = 1.602$ .
  - a) Queremos saber si se cumplen los requisitos de la garantía, o sea, si la media de la variable transformada es menor que  $\log_{10}(100) = 2$ . Utiliza el contraste adecuado para responder la pregunta anterior, usando  $\alpha = 0.05$ .
  - b) Para reducir el coste, se prueba el telemantenimiento con un sistema alternativo más barato, calculando los logaritmos de los tiempos de reparación con este sistema en las MISMAS 4 averías anteriores:  $\log_{10}(97) = 1.987$ ,  $\log_{10}(36) = 1.556$ ,  $\log_{10}(60) = 1.778$ ,  $\log_{10}(39) = 1.591$ , respectivamente. Utiliza el contraste adecuado para comprobar si el sistema alternativo reduce el (logaritmo del) tiempo de reparación, usando  $\alpha = 0.05$ .
  
4. (2 puntos) Una fábrica de papel utiliza gráficos de control para monitorizar las imperfecciones en rollos de papel.
  - a) Se inspeccionan 20 muestras preliminares de 10 rollos cada una, recogiendo el número de imperfecciones totales: 6 7 10 9 8 6 5 17 6 15 4 7 3 5 7 6 8 9 10 13.
    - i) Utilizando todos los datos, calcula los límites de control para una gráfica U.
    - ii) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - b) Si en el proceso anterior, consideráramos como característica de calidad a controlar el grosor del papel, ¿qué tipo de gráficas usarías? ¿por qué?
  - c) Si ahora consideráramos como defectuosos aquellos rollos con alguna imperfección y controláramos el número de rollos defectuosos de los 10 seleccionados, ¿qué tipo de gráficas usarías? ¿por qué?
  - d) Supongamos que los papeles deben cortarse según cierta forma y que se controla la longitud entre dos puntos determinados, tomando cada vez muestras de tamaño 7. Supongamos que la media y desviación típica del proceso bajo control fuera respectivamente: 2.05 y 0.3, y los límites del gráfico  $\bar{X}$ : 1.71 y 2.39. Si las especificaciones son  $2 \pm 0.1$ .
    - i) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.  
Supongamos que la media se corre a 1.7.
    - ii) ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento? ¿Cuál es la ARL después del corrimiento?

Examen 25/1/2006. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado. MODELO B.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

- (1.75 puntos) Estamos interesados en analizar la productividad media por hora de tres técnicas de montaje distintas de un cierto mecanismo: A, B o C. Se tiene la sospecha de que el día de la semana puede influir en la productividad, de forma que se han tomado datos para cada procedimiento de forma aleatoria para cada día: lunes, martes, miércoles y jueves. Los datos aparecen a continuación:

Procedimiento	Día semana			
	Lu	Ma	Mi	Ju
A	2.6	2.1	3.5	2.6
B	3.4	2.7	3.9	3.2
C	2.7	2.1	3.1	2.6

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas de productividad según el método utilizado.

Analysis of Variance for Prodhora

Fuente	Suma Cuadrados	GL	Media Cuadrados	F
-----				
A:Procedimiento				
B:Día	2,1825			
RESIDUAL			0,0141667	
-----				
TOTAL	3,3625			
-----				

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de  $X$ 's.
- (2 puntos) Se desea comprobar si un cambio en un determinado proceso industrial es beneficioso, si mejora el anterior. Para ello se toman dos muestras, antes y después del cambio. Se observa que 63 de 270 elementos son defectuosos en el primer caso, mientras que una vez implementado el cambio, obtenemos 45 defectuosos de 300.
    - Utiliza el contraste adecuado para comprobar si el cambio mejora el proceso anterior, a nivel de significación de 0.05.
    - Considera sólo el proceso tras el cambio. Queremos estimar la proporción de defectuosos, deseando estar al menos 95 % seguros que el error es como mucho de 0.03.
      - ¿Cómo ha de ser de grande la muestra si usamos la proporción anterior como una muestra preliminar que nos proporciona una primera estimación?
      - ¿y si no tuviéramos idea de cuál pueda ser la proporción real?

3. (1 punto) Consideremos 3 proveedores  $A$ ,  $B$  y  $C$ ; 200 componentes de cada proveedor son aleatoriamente seleccionadas para examinarlas obteniéndose los siguientes resultados:

Calidad	Proveedor		
	A	B	C
Defectuoso	4	6	16
No defectuoso	196	194	184

¿La proporción de defectuosos es la misma para los 3 proveedores, o sea, son homogéneas las poblaciones? (considera  $\alpha = 0.01$ ).

4. (0.75 puntos) Se desea comprobar dos algoritmos de inversión de grandes matrices. Como el tiempo de resolución depende de las características de cada matriz, se invirtieron las mismas matrices con los dos algoritmos, con los siguientes tiempos de resolución:

Matrices	Algoritmo 1	Algoritmo 2
M1	5.6	5.4
M2	3.8	3.8
M3	3.9	4.7
M4	2.3	2.1
M5	1.2	1.7
M6	4.1	4.2

Determina razonadamente si existen diferencias entre las velocidades medias de ambos algoritmos, estableciendo el intervalo de confianza apropiado al 95 %.

5. (2 puntos) Una fábrica de papel utiliza gráficos de control para monitorizar las imperfecciones en rollos de papel.
- Se inspeccionan 20 muestras preliminares de 10 rollos cada una, recogiendo el número de imperfecciones totales: 5 17 6 15 4 7 3 5 7 6 8 9 10 13 6 7 10 9 8 6.
    - i) Utilizando todos los datos, calcula los límites de control para una gráfica U.
    - ii) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - Si en el proceso anterior consideráramos como defectuosos aquellos rollos con alguna imperfección y controláramos el número de rollos defectuosos de los 10 seleccionados, ¿qué tipo de gráficas usarías? ¿por qué?
  - Si ahora consideráramos como característica de calidad a controlar el grosor del papel, ¿qué tipo de gráficas usarías? ¿por qué?
  - Supongamos que los papeles deben cortarse según cierta forma y que se controla la longitud entre dos puntos determinados, tomando cada vez muestras de tamaño 7. Supongamos que la media y desviación típica del proceso bajo control fuera respectivamente: 2.05 y 0.3, y los límites del gráfico  $\bar{X}$ : 1.71 y 2.39. Si las especificaciones son  $2 \pm 0.1$ .
    - i) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.  
Supongamos que la media se corre a 1.85.
    - ii) ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento? ¿Cuál es la ARL después del corrimiento?

Examen 14/9/2006. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluidas las del enunciado.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

- (1.75 puntos) La rapidez de desplazamiento en la pantalla es una consideración importante en el desarrollo de tarjetas gráficas a color. Se realiza un estudio para comparar el tiempo, en segundos, para desplazarse una pantalla en diversos documentos de *Macrosoft Word* con cuatro tarjetas gráficas distintas en monitores de 24 pulg. La prueba emprendida es la de rendimiento Hydra Quick Draw. Los datos obtenidos (basado en información de "Gauging Video Speed", MAC WORLD, junio 1993, p. 28) fueron:

Tarjetas	Observaciones				
A	79	85.3	86.2	82	81.7
B	48.3	42.1	43.5	40.6	38.6
C	79.2	84.7	85	88.2	76.3
D	51.6	59.4	57.3	59	58.7

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas según las tarjetas.

ANOVA Table for Tiempo by Tarjeta

Fuente	Suma Cuadrados	GL	Media Cuadrados	F
Tratamientos			1975,71	
Error	222,088			
Total				

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. Según estos resultados, ¿qué tarjeta recomendarías?
- (1.75 puntos) El tiempo en segundos necesario para conectarse con Internet mediante un servicio con el que debe marcarse un número, recibe influencia de diversos factores, como  $n^\circ$  de líneas telefónicas disponibles en el área de llamadas locales, hora del día, día de la semana,  $n^\circ$  de usuarios en el área, etc. En un área dada, se recogieron el mismo día de la semana 21 datos en dos momentos del día, mañana (9-11 a.m.) y noche (10-12 p.m.), resultando con medias respectivas: 45 y 33; y desviaciones típicas respectivas: 15 y 11.
    - Comprueba razonadamente si las varianzas difieren, calculando el intervalo de confianza al 95 % adecuado (asume normalidad).
    - Comprueba razonadamente si los tiempos medios difieren, calculando el intervalo de confianza al 99 % adecuado. (Si no has resuelto el apartado a), asume varianzas iguales).
    - Utilizando los mismos datos del enunciado, ¿cómo podría obtenerse un intervalo más estrecho que el del apartado b)?
  - (1 punto) Los sistemas de cómputo se colapsan por muchas razones. Se piensa que el 10 % de los colapsos se debe a fallos de software; 5 % a fallos de hardware; 25 % a errores

del operador; 40% a sobrecargas del sistema, y el resto, a otras causas. En un estudio prolongado, se observan 150 colapsos de computadora, y cada una se clasifica según su causa probable, determinándose que: 13 resultan de fallos de software; 10 de fallos de hardware; 42 de errores del operador; 65 de sobrecargas del sistema, y el resto de otras causas. Verifica si estos datos se ajustan a la distribución de partida (considera  $\alpha = 0.05$ ).

4. (1 punto) Se ha desarrollado un nuevo algoritmo para solucionar problemas objetivos múltiples de cero-uno en programación lineal. Se piensa que el nuevo algoritmo soluciona los problemas con mayor rapidez que el algoritmo en uso. Seleccionamos al azar varios problemas, que se resuelven con ambos algoritmos, con los siguientes tiempos de resolución:

Problemas	Algoritmo vigente	Algoritmo nuevo
P1	8	0.7
P2	24.7	0.8
P3	28.3	0.8
P4	9	0.7
P5	25	0.8
P6	14	0.8

Usa el contraste que determine si el algoritmo nuevo mejora el vigente (con  $\alpha = 0.05$ ).

5. (2 puntos) En un proceso de empaquetado de sal, se está preocupado por el contenido de humedad en los paquetes, pues la variación en la humedad del ambiente de la planta puede variar el contenido de humedad promedio de los paquetes. Por ello, se extraen 5 paquetes cada 15 minutos durante 3 horas, midiéndose el contenido de humedad en cada paquete. Los datos de las medias y rangos para las 5 mediciones tomadas cada vez, son:

Medias: 2.3, 2.5, 2.4, 2.3, 2.5, 2.4, 2.7, 2.9, 3.4, 2.8, 2.7, 2.6

Rangos: 0.8, 0.4, 0.6, 0.4, 0.6, 0.5, 0.4, 1.4, 0.6, 0.7, 0.5, 0.6

- Utilizando todos los datos, calcula los límites de control para las gráficas de  $\bar{X}$  y  $R$ .
- ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- Tras realizar b), estima la media y desviación típica del proceso.
- Supongamos que la media se corre a 2.7. ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento? ¿Cuál es la ARL después del corrimiento?
- Si en el proceso anterior, consideráramos como característica de calidad a controlar simplemente si se supera o no un cierto contenido de humedad, ¿qué tipo de gráficas usarías? ¿por qué?
- En la planta se ha decidido instalar un deshumidificador para estabilizar la humedad del ambiente, tras lo cual (elige la opción correcta, razonándolo):
  - el seguimiento del proceso debe seguir realizándose constantemente.
  - el seguimiento debe detenerse durante cierto tiempo, ya que es poco probable que el proceso se salga fuera de control inmediatamente.
  - no debe darse seguimiento nuevamente al proceso, pues está bajo control.

Examen 24/1/2007. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas, incluidas las del enunciado. MODELO A.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

1. (2 puntos) Un sistema informático de un banco monitoriza el tiempo que se tarda en hacer las transmisiones entre dos nodos de una red en la que se precisa de una fiabilidad muy alta. De esta forma, se pueden detectar anomalías y evitar que el sistema falle y se produzca un bloqueo que interrumpa el servicio, lo que provocaría cuantiosas pérdidas. Para hacer dicha monitorización, el sistema toma conjuntos de 4 transmisiones consecutivas periódicamente cada minuto y calcula la media y el rango de las duraciones. En condiciones normales de transmisión, el tiempo que se tarda en hacer la transmisión sigue una distribución normal. Las medias y rangos para 15 muestras son los siguientes (en segundos):

Medias: 0.22; 0.14; 0.18; 0.17; 0.29; 0.17; 0.19; 0.16; 0.21; 0.22; 0.2; 0.2; 0.15; 0.18; 0.23

Rangos: 0.1; 0.05; 0.08; 0.09; 0.07; 0.06 ; 0.1; 0.19; 0.07; 0.08; 0.05; 0.08; 0.07; 0.04; 0.07

- a) Utilizando todos los datos, calcula los límites de control para las gráficas de  $\bar{X}$  y  $R$ .
- b) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- c) Tras realizar b), estima la media y desviación típica del proceso.
- d) Si el tiempo medio de la transmisión aumentara en 0.04 segundos, ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el desajuste? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 0.2 y 0.05, respectivamente).
- e) Si quisieramos reducir el tiempo que se tardará en detectar dichos desajustes, ¿qué estrategias se podrían seguir?
2. (2.75 puntos) A fin de verificar la adecuación de un sistema informático interactivo de venta de entradas de cine, se controla el tiempo de servicio de los usuarios. Para que este sistema sea satisfactorio, el tiempo de servicio medio por cliente no debe superar los 2 minutos. En efecto, los estudios realizados mostraron que un tiempo medio superior produce unas colas demasiado largas, y una espera que el usuario no está dispuesto a soportar; por lo tanto, el cine perderá clientes y dinero si el requisito mencionado no se satisface. Para controlar el tiempo de servicio, se observa una muestra aleatoria simple de 31 usuarios en uno de los cines de la cadena (en el "ABC vamos ar sine"), para saber si se debe o no proceder a la modificación del sistema informático de venta.
- a) Utiliza el contraste adecuado para comprobar si se debe modificar el sistema informático actual, a nivel de significación de 0.05, sabiendo que el tiempo de servicio medio observado en la muestra es de 2.17 minutos y la desviación típica de 0.4 minutos.
- b) En dicho cine de la cadena, se instala de prueba un nuevo sistema informático, por comprobar si mejora el anterior y recomendar su cambio en el resto de cines de la cadena con largas colas. Esta vez, sólo se observan 11 usuarios, con media y desviación típica de 1.9 y 0.35, respectivamente. Asumiendo normalidad:
- i) Comprueba razonadamente si las varianzas de los tiempos de servicio difieren, calculando el intervalo de confianza al 95% adecuado.
  - ii) Usa el contraste que determine si el sistema nuevo mejora el anterior, en cuanto a los tiempos de servicio medios se refiere (con  $\alpha = 0.05$ ).

- c) Centrandonos únicamente en el sistema nuevo, si deseamos que el error en la estimación del tiempo de servicio medio sea inferior a 0.05 minutos con una confianza del 95 %, y teniendo en cuenta que podríamos asumir  $\sigma = 0.3$ , ¿cuántos usuarios deberían observarse como mínimo?
3. (1.75 puntos) Tres algoritmos (A, B, C) de búsqueda de caminos mínimos han sido probados sobre los mismos 6 grafos. Los tiempos de CPU obtenidos han sido:

Algoritmo	Grafos					
	1	2	3	4	5	6
A	275	230	267	282	178	222
B	258	234	280	286	218	234
C	340	325	314	320	310	298

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas según el algoritmo utilizado.

Análisis de Varianza para Tiempo

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
-----				
A: Algoritmo				
B: Grafos			1801,79	
RESIDUAL	3740,56			
-----				
TOTAL	33080,3			
-----				

- a) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- b) Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's.
4. (1 punto) Se desean comparar cuatro algoritmos de inversión de matrices en ciertas condiciones extremas en cuanto al tamaño de la matriz. Para ello, se ejecuta cada algoritmo 100 veces en diferentes matrices, de gran dimensión, generadas aleatoriamente. Se anota el resultado de cada ejecución según se produzca: a) un error irrecuperable, b) un error recuperable, y c) concluya sin errores. Se desea saber si los cuatro algoritmos tienen el mismo perfil de errores, es decir, se desea contrastar su homogeneidad. (Considera  $\alpha = 0.01$ ).

Algoritmo	Tipo de Resultado		
	Irrecuperable	Recuperable	Sin error
A	16	19	65
B	18	27	55
C	14	23	63
D	23	36	41

Examen 3/9/2007. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas, incluidas las del enunciado.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

1. (1.75 puntos) La compañía Alexa (www.alexa.com) es una empresa que, entre otras actividades, monitoriza el número de visitas que reciben las diferentes páginas web. Consideremos el número de visitas que, según Alexa, ha recibido la web de la UJI en los últimos 105 días y que los agrupamos en subgrupos de 7 días, obteniéndose un total de 15 muestras. Suponiendo que el número de visitas sigue una distribución normal, y que las medias y rangos para 15 muestras son los siguientes:  
Medias: 5000; 4950; 5080; 4990; 5070; 5000; 5120; 5200; 5040; 4960; 5000; 5020; 4970; 5050; 4980  
Rangos: 180; 200; 170; 150; 220; 190; 200; 230; 190; 170; 220; 250; 180; 210; 200
  - a) Utilizando todos los datos, calcula los límites de control para las gráficas de  $\bar{X}$  y  $R$ .
  - b) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - c) Tras realizar b), estima la media y desviación típica del proceso.
  - d) Si el número medio de visitas bajase en 100, ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el cambio? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 5000 y 70, respectivamente).
  
2. (2.75 puntos) Se han propuesto diversos métodos para detectar si una persona no autorizada (un intruso) accede a una cuenta con el nombre de usuario y contraseña (robada o descifrada) correctas. Uno de ellos consiste en medir el tiempo entre las pulsaciones de las teclas. A un importante usuario autorizado (debidamente identificado) se le ha medido dicho tiempo, dando las 121 observaciones recogidas una media de 0.2 segundos y una desviación típica de 0.07 segundos. En base a estos datos:
  - a) Construye un intervalo de confianza al 99% para el tiempo medio entre pulsaciones de dicho usuario.
  - b) Si deseamos que el error en la estimación del tiempo medio anterior sea inferior a 0.01 segundos con una confianza del 95%, y teniendo en cuenta que podríamos asumir  $\sigma = 0.07$ , ¿cuántas observaciones deberían recogerse como mínimo?
  - c) Un día se graban los siguientes tiempos entre pulsaciones para una entrada correcta de dicho usuario en segundos: 0.38, 0.31, 0.24, 0.2, 0.31, 0.34, 0.42, 0.09, 0.18, 0.46, 0.21. Asumiendo normalidad:
    - i) Comprueba razonadamente si las varianzas de los tiempos entre pulsaciones difieren, calculando el intervalo de confianza al 95% adecuado (usa los grados de libertad más cercanos de la tabla apropiada).
    - ii) En cuanto a los tiempos medios entre pulsaciones se refiere, usa el contraste que determine si hay evidencia de que un intruso ha accedido a la cuenta de dicho usuario (con  $\alpha = 0.05$ ).

3. (2 puntos) En una región de Egipto, se recogió una muestra de calaveras de cinco épocas distintas a las que se les midió su amplitud máxima, con el fin de estudiar si hay diferencia en las medidas de acuerdo a las distintas épocas, lo cual indicaría que se produjo un mestizaje con poblaciones inmigrantes. Aunque el estudio fue mayor, consideremos sólo las siguientes observaciones:

Época	Observaciones			
4000 A.C.	131	125	131	124
3300 A.C.	124	130	133	128
1850 A.C.	137	129	140	136
200 A.C.	137	138	135	140
150 D.C.	132	140	147	136

A partir de estos datos se ha construido la siguiente tabla ANOVA:

Análisis de Varianza para Amplitud máxima

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Tratamientos			103,075	
Error				
TOTAL	698,55			

- a) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
  - b) Calcula la LSD al 95% y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Hubo algún período de años clave?
  - c) Si en el estudio se considera también otra región de Egipto, en la que también se recogen calaveras de las distintas épocas, ¿qué tipo de diseño experimental emplearías?
4. (1 punto) Se ha hecho un seguimiento durante una serie de días del número de mensajes spam al día que un cierto usuario recibe en su correo electrónico. En base a dichos datos, que se recogen en la tabla siguiente, decide si se ajustaría a una distribución de Poisson, considerando  $\alpha = 0.05$ .

Número spam diario	0	1	2	3	4	5	$\geq 6$
Frecuencia observada	35	42	55	40	15	10	3

Examen 2/2/2008. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas, incluidas las del enunciado. MODELO A.

En tus cálculos emplea más de 4 decimales, y haz la letra clara. ENTREGA EL ENUNCIADO.

1. (1.5 puntos) Una agencia de investigación agraria está probando cuatro métodos de riego para naranjos, para ver si son más o menos rentables con igual cantidad de agua gastada. El método 1 es por goteo con válvula reguladora, el método 2 por goteo sin válvula reguladora, el método 3 es por aspersión, y el método 4 es el tradicional de regueros. Se han escogido tres poblaciones para hacer las pruebas, y en cada una de ellas se han probado cuatro parcelas. Los rendimientos obtenidos en Kg/árbol fueron:

Método	Poblaciones		
	Burriana	Almenara	Puzol
1	51.3	44.8	42.5
2	42.2	40.1	39.3
3	43.5	40.5	39.6
4	45.9	42.3	39.9

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre los métodos de riego.

Análisis de Varianza para Kg/árbol

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Metodo				
Poblacion		30,773		
RESIDUAL	11,180			
TOTAL	130,29			

- a) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- b) Calcula la LSD al 95% y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Cuál/es método/s serían mejores?
2. (3 puntos; 3.5 sin Problemas Intentados) La empresa de desarrollo de software *UJSoftware* distribuye una aplicación para procesar estadísticas de bases de datos de gran volumen. La complejidad de la aplicación hace que a veces en tiempo de ejecución, se produzcan errores. Una ejecución puede ser de uno de los siguientes tipos: 1) Limpia: sin errores, 2) Defectuosa: con *warnings*, pero finaliza normalmente, o 3) Incompleta: la ejecución finalizada de forma súbita (error grave). Se pretende revisar la última versión del programa, y para ello se dispone de dos opciones: a) un programador que prepara de forma manual las bases de datos y ejecuta la aplicación, registrando los tipos de ejecuciones; b) un programa para testar de forma automática las ejecuciones. Este programa puede generar las bases de datos y registra los resultados de las ejecuciones. En todas las pruebas realizadas, las bases de datos son equivalentes. Responde a los objetivos planteados:
- a) Primero, se quiere comprobar si las dos opciones para verificar el programa son equivalentes, o sea, la homogeneidad ( $\alpha=0.05$ ). Para ello, de forma manual se han generado 50 bases de datos, y se ha ejecutado manualmente la aplicación, contando los diferentes tipos de ejecución registrados. Por otra parte, el programa para testar de forma automática ha generado otras 200 bases de datos y ha contado también los tipos de ejecución. Los resultados son los siguientes:

Opción	Tipo de Ejecución		
	Limpia	Defectuosa	Incompleta
Manual	24	13	13
Automática	156	29	15

- b) En segundo lugar, se quiere comprobar (contrastar) si el proceso manual en promedio, es capaz de detectar más *warnings* que el proceso automático con  $\alpha=0.05$ . Para ello, se generan 61 bases de datos de prueba y cada una de ellas se ejecuta simultáneamente con los dos procedimientos, manual y automático. Para cada una de estas bases de datos, se cuenta el número de *warnings* detectados por los dos procedimientos, que asumiremos Normales. El resumen descriptivo de la muestra obtenida es el siguiente (la variable 'Manual - Automática' es la resta de ambas):

Variable	N	Media	Mediana	Desviación típica
Manual	61	2,800	3,000	2,609
Automática	61	2,267	2,000	1,885
Manual - Automática	61	0,533	1,000	0,999

- c) Si contáramos con otro procedimiento automático, además de las dos opciones anteriores, explica cómo se podría comprobar si difieren en promedio en detección de *warnings*, como en el apartado anterior.
- d) Tercero, se analizó el tiempo que invirtieron dos de sus programadores en resolver 21 *warnings* diferentes para cada programador, pero similares entre ellos, pues se piensa que no son igualmente sistemáticos, es decir, que difieren en variabilidad. Los datos aparecen a continuación. Utiliza el intervalo de confianza al 95% adecuado, para comprobar lo anterior razonadamente, y asume normalidad.

Tiempo (en horas)	N	Media	Mediana	Desviación típica
Tiempo invertido por el programador 1	21	2,798	2,534	0,820
Tiempo invertido por el programador 2	21	2,593	2,575	0,603
Tiempo invertido por el programador 1 - 2	21	0,205	0,197	0,347

3. (2.5 puntos) Se desea construir un gráfico de control para controlar un proceso que fabrica diodos para un circuito impreso. Se tienen 20 muestras, cada una formada por 50 diodos.
- a) El número de diodos defectuosos en cada una de las muestras aparece a continuación: 4 5 3 1 4 5 7 5 10 6 5 1 3 1 2 3 5 4 6 2.
- i) Utilizando todos los datos, calcula los límites de control para un gráfico de control apropiado.
  - ii) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- b) Al cabo del tiempo, tras un reajuste, el operario sospecha que la proporción de diodos defectuosos puede haber disminuido, así que toma aleatoriamente 200 diodos de la cadena de producción y comprueba que 10 son defectuosos. Utiliza el contraste adecuado para comprobar la sospecha del operario, a nivel de significación de 0.05. (Si no has realizado el apartado anterior, usa los datos del enunciado.)
- c) Si deseamos que el error en la estimación de la proporción de defectuosos tras el reajuste, sea inferior a 0.02 con una confianza del 95%, y usando la proporción del apartado anterior como una muestra preliminar que nos proporciona una primera estimación, ¿cuál será el tamaño muestral requerido?
- d) En otro departamento, están interesados en el proceso de manufactura de unos aparatos electrónicos, cuya corriente tiene especificaciones de  $100 \pm 10$  miliamperes. La media y desviación típica del proceso son 108 y 1.5, respectivamente, y se miden 6 aparatos cada vez.
- i) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.
  - ii) Calcula la probabilidad de que el proceso produzca corrientes fuera de los límites de las especificaciones.

Examen 1/9/2008. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR.  
INDICA tus cálculos, emplea más de 4 decimales, y haz la letra clara CON BOLÍGRAFO.

1. (1.75 puntos) Este verano ha aparecido en los medios de comunicación un estudio realizado sobre los beneficios del *Brain Training*. En concreto, durante quince semanas un grupo de personas jugó diariamente durante un periodo corto de tiempo. A continuación, aparece un subconjunto adaptado de estos datos, con los tiempos en segundos (que asumiremos normal) para el ejercicio de agilidad mental, Cálculo 20, en la primera y última semana para 7 individuos:

	Semana 1	Semana 15
Sujeto 1	60	31
Sujeto 2	50	25
Sujeto 3	54	20
Sujeto 4	74	35
Sujeto 5	58	24
Sujeto 6	65	23
Sujeto 7	57	22

- a) Utiliza el contraste de hipótesis adecuado para determinar en base a dichos datos, si el uso del *Brain Training* es beneficioso (usa  $\alpha = 0.05$ ).
- b) Centrándonos únicamente en las mediciones para la semana 15:
- i) Calcula el intervalo de confianza al 95% para el tiempo medio invertido en Cálculo 20 tras el entrenamiento de las 15 semanas.
- ii) Si deseamos que el error en la estimación del tiempo medio anterior sea inferior a 1.5 segundos con una confianza del 99%, y estimando  $\sigma$  con la desviación típica de los datos anteriores en la semana 15, ¿cuántos sujetos tendrían que considerarse?
2. Antes de sacar al mercado un cierto software sobre edición de vídeo, se realiza un test de utilización (usability testing), en el que potenciales usuarios prueban el producto y se recogen sus datos, para con ellos refinar el producto antes de sacarlo a la venta.

- a) (0.75 puntos; 1.25 sin problemas intentados) Entre los datos recogidos estuvo el número de llamadas al 'help-desk' en un hora. En base a dichos datos, recogidos en la tabla siguiente, decide si se ajustaría a una distribución de Poisson, con  $\alpha = 0.05$ .

Número de veces que se acude a la ayuda	0	1	2	3	4	5	$\geq 6$
Frecuencia observada	13	16	14	6	2	3	1

- b) (1.5 puntos) En otro punto del estudio, se quería comparar 3 distintas interfaces. Para ello, a 5 sujetos se les midió el tiempo empleado en completar una determinada tarea para cada una de las interfaces, cuyo orden de presentación fue aleatoriamente seleccionado. A continuación, aparecen estos datos:

Interfaz	Sujetos				
	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5
1	51	49	46	43	45
2	63	61	53	60	59
3	47	44	43	45	44

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre las interfaces.

Análisis de Varianza

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Interfaz				
Sujetos			17,2667	
RESIDUAL	37,7333			
TOTAL	726,4			

- i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- ii) Calcula la LSD al 95% y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Cuál/es interfaz/s serían mejores?

3. (1.25 puntos) Desde hace unos años, en los grandes torneos de tenis se viene aplicando un sistema llamado 'Ojo de halcón', que consiste en una serie de cámaras dispuestas en la pista, y con cuya información se reconstruye, mediante algoritmos matemáticos, la trayectoria más probable de la pelota. Además de servir de ayuda en las decisiones de los jueces, también almacena los datos pormenorizados que pueden ser analizados estadísticamente. Aquí nos vamos a restringir, sin embargo, a datos más fácilmente accesibles, como es la efectividad de los primeros servicios en las finales de Roland Garros y Wimbledon en 2008, entre Nadal y Federer:

Roland Garros		Wimbledon	
Nadal	Federer	Nadal	Federer
50 de 67	53 de 77	159 de 218	128 de 195

- a) Mediante el intervalo de confianza apropiado con  $\alpha = 0.1$ , determina si existe diferencia en cuanto a la efectividad entre Nadal y Federer este año (por tanto, considera conjuntamente ambos torneos).
  - b) Centrándonos ahora en Nadal, si queremos que la estimación sobre su efectividad real en sus primeros servicios en 2008, sea inferior a 0.03 con una confianza del 90%, y usando los anteriores datos como preliminares, ¿cuál será el tamaño muestral requerido?
4. (1.75 puntos) Se muestra a continuación las medias y rangos para 15 días, cada uno basado en 6 observaciones diarias del índice de refracción de un cable de fibra óptica.

Medias: 95.7; 95.4; 96.6; 97.4; 96.9; 96.8; 96.5; 98.3; 96; 97.2; 96.5; 96.6; 96.4; 95.5; 97.4

Rangos: 3.2; 6.4; 3.6; 3.2; 1.9; 3.3; 3.4; 3.5; 3.1; 2.3; 3.1; 1.4; 3.8; 1.5; 3.4

- a) Utilizando todos los datos, calcula los límites de control para las gráficas de  $\bar{X}$  y  $R$ .
- b) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- c) Tras realizar b), estima la media y desviación típica del proceso.
- d) Si el índice de refracción medio bajase en 0.5, ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el cambio? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 97 y 1.3, respectivamente).

Examen 21/1/2009. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG.

Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR. MODELO A.  
INDICA tus cálculos, emplea más de 4 decimales, y haz la letra clara CON BOLÍGRAFO.

1. (1.75 pts) Un capítulo de un programa de TV de divulgación científica trató el comportamiento humano:
- a) Un economista conductual llevó a cabo el siguiente experimento: se ofrecen dos tipos de chocolates a distintas personas para ver por cuál optarían en función de la calidad-precio. Uno es un chocolate belga de una marca muy prestigiosa de buena calidad cuyo precio real es 0,60 €/unidad. El otro es un chocolate de una marca de supermercado de menor calidad cuyo precio real es 0,20€/unidad. En el primer caso se ofrecen los chocolates con los siguientes precios: el chocolate belga a 0,15€y del supermercado a 0,01€. De 100 personas, 73 compran el chocolate belga, y 27 el chocolate del supermercado. En el segundo caso se ofrecen los chocolates con el mismo margen de precio, pero se decide hacer una rebaja de 0,01€, con lo que los precios serían: chocolate belga a 0,14€y chocolate de supermercado GRATIS (0€). Ahora, 31 personas optan por el chocolate belga y 69 por el del supermercado. Mediante el intervalo de confianza al 95 % adecuado, explica razonadamente si hay diferencias en las elecciones del chocolate de supermercado.
  - b) Piensa en el desastre del Titanic como si se tratara de un experimento para averiguar cómo se comportó la gente de aquel tiempo al enfrentarse con la muerte en una situación en la que sólo algunos podían salvarse. Los pasajeros son una muestra de la población de su nivel económico. Según la tabla siguiente, y fijándonos SÓLO en los hombres, ¿la proporción de supervivientes es la misma según el nivel económico, o sea, son homogéneas sus distribuciones? (Usa  $\alpha = 0.05$ ).

Nivel económico	Hombres		Mujeres	
	Murieron	Sobrevivieron	Murieron	Sobrevivieron
Alto	111	61	6	126
Medio	150	22	13	90
Bajo	419	85	107	101

2. Uno de los problemas propuestos para intentar en el curso, versaba sobre un problema sobre simulación. Imagina que disponemos de un programa para simular el sistema y obtener los resultados (asumidos normales), en los dos contextos siguientes:
- a) (2.25 puntos; 2.75 sin Problemas Intentados) Supermercado: se plantean dos estrategias de distribución de cajas de pago y colas. Para cada una de las cuales, hemos simulado réplicas independientes y obtenido el tiempo medio que los clientes estarían haciendo cola:  
Estrategia 1 (la actualmente en uso): 1.91, 1.82, 1.71, 1.83, 2.2, 2.4  
Estrategia 2: 1.53, 1.66, 1.24, 2.34, 2
  - i) Calcula un intervalo de confianza al 95 % para el cociente de varianzas y determina (razonando porqué) si existe diferencia entre ellas.
  - ii) ¿Puede concluirse al nivel  $\alpha = 0.05$  que la estrategia 2 es mejor que la actual en cuanto a tiempos medios de espera se refiere? Utiliza el contraste adecuado para responder la pregunta anterior. (Si no has resuelto el apartado i), asume varianzas distintas).
  - iii) Según el resultado del apartado anterior (sino lo has hecho, usa uno que pienses coherente), indica razonadamente cuánto puede valer el p-valor.
  - iv) Centrándonos únicamente en la estrategia 2, si deseamos que el error en la estimación del tiempo de espera medio sea inferior a 0.1 minutos con una confianza del 95 %, y teniendo en cuenta que podríamos asumir  $\sigma = 0.35$ , ¿cuántas réplicas deberían realizarse como mínimo?
  - b) (1.5 pts) Control de la producción: en la tabla siguiente aparecen los costes (gastos) totales medios por mes para cinco réplicas independientes de 4 políticas de inventario diferentes:

Política	Observaciones				
Política 1: (20,40)	126.9	124.3	126.7	122.6	127.3
Política 2: (20,80)	118.2	120.2	122.4	122.7	119.4
Política 3: (40,60)	120.7	129.3	120.6	123.6	127.3
Política 4: (40,100)	131.6	137	129.9	129.9	131

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre las políticas.

**Análisis de Varianza**

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Política				
Error			7,9705	
TOTAL	459,172			

- Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- Calcula la LSD al 95% y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Cuál/es política/s serían mejores?

3. (1.5 pts) Para controlar la calidad del agua, se seleccionan cada semana 5 muestras de agua de una fuente de abastecimiento, y se determina la concentración promedio de una sustancia tóxica. Para doce semanas estos son los datos en partes por millón (ppm):

$$\sum_{i=1}^{12} \bar{x}_i = 61'8, \text{ y } \sum_{i=1}^{12} r_i = 13'95.$$

- Encuentra los límites de control de los gráficos  $\bar{X}$  y  $R$ .
- Suponiendo que el proceso está bajo control, estima la media y desviación típica del proceso. Si se considera peligrosa una concentración superior a 6 ppm,
- ¿cuál es la probabilidad de que una medición del agua de la fuente sea considerada peligrosa?
- Es importante detectar rápidamente en el gráfico  $\bar{X}$ , un desplazamiento de la concentración media a 5.25 ppm, ¿cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento?
- ¿Cuál sería el número medio de muestras a representar hasta detectar el desplazamiento? ¿qué estrategias podrían seguirse para disminuir el tiempo de detección?

Examen 16/9/2009. IG23 AMPLIACION DE ESTADISTICA. ITIG.

Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR.  
INDICA tus calculos, emplea mas de 4 decimales, y haz la letra clara CON BOLIGRAFO.

1. (1.5 pts; 2 sin problemas intentados) Un servidor web de una red informatica recibe un 0.5 % de intentos de ataque por virus de cada intento de conexion. La organizacion que posee dicho servidor hace una campaña de publicidad para difundir la direccion de su sitio web. Al día siguiente de dicha campaña, el servidor recibe 830 peticiones de conexion, de las que 10 son ataques por virus.
  - a) ¿Hay evidencia suficiente para afirmar que la campaña publicitaria ha hecho aumentar la proporción de conexiones que son debidas a virus? ( $\alpha = 0.05$ ).
  - b) Para disminuir los ataques por virus, se coloca un filtro que desvía los mensajes sospechosos para ser analizados. Al día siguiente de poner el filtro se tiene que de los 700 mensajes que pasaron el filtro (por no detectarse indicios de sospecha alguna), 5 resultaron ser ataques por virus. ¿Puede afirmarse (con  $\alpha = 0.05$ ) que el filtro ha mejorado la situación respecto a la anterior (tomar como situación anterior, la que se tiene tras la campaña publicitaria, donde con 830 conexiones hubo 10 ataques)?
2. (1 pt.) En un servicio de informatica se han recogido los datos sobre las 3 marcas de monitores con que trabajan y si necesitaron reparacion en garantía. Determina si son independientes estas dos variables, con  $\alpha = 0.05$ .

Reparacion	Marca		
	A	B	C
S	268	423	418
No	1168	1223	1345

3. (1.5 puntos) Se quiere estudiar si es efectiva una campaña de información sobre el uso de la batería del portátil. Se solicitó el teléfono de contacto de una serie de clientes (aleatoriamente asignados) que adquirieron un modelo determinado. A la mitad de ellos (también de forma aleatoria divididos) se les envió un folleto con la información. Al cabo de un año, se contactó con los clientes, preguntándoles sobre la duración usual (que consideraremos Normal) que obtienen de la batería. Se consiguieron datos de 12 clientes de los que tenían folleto, y 8 de los que no lo tenían. La duración media para el grupo con folleto fue de 2.5h. y la desviación típica de 0.21h., mientras que para el otro grupo, fueron de 2.1h. y 0.24h respectivamente.
  - a) Mediante el intervalo de confianza apropiado con  $\alpha = 0.1$ , determina si la variación en la duración en ambos grupos difiere.
  - b) Mediante el intervalo de confianza apropiado con  $\alpha = 0.1$ , determina si la duración promedio de ambos grupos difiere.
  - c) Centrandonos unicamente en los clientes con folleto, si deseamos que el error en la estimación de la duración media sea inferior a 0.06h con una confianza del 95 %, y teniendo en cuenta que podremos asumir  $\sigma = 0.21$ , ¿cuántas muestras deberán recopilarse?
4. (1.5 puntos) Se quiere estudiar el efecto de tres diseños para un blog (distintos colores y letra) en la velocidad de lectura. Para ello, a 5 sujetos se les midió el tiempo empleado en leer en voz alta

ciertos bloques del blog con cada diseño, cuyo orden de presentacion fue aleatoriamente seleccionado. A continuacion, aparecen estos datos:

Diseño	Sujetos				
	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5
1	51	49	46	43	45
2	63	61	53	60	59
3	47	44	43	45	44

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre los diseños.

#### Análisis de Varianza

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Diseño				
Sujetos			17,2667	
RESIDUAL	37,7333			
TOTAL	726,4			

- i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba ( $\alpha = 0.05$ ), explicando el resultado que obtengas.
  - ii) Calcula la LSD al 95% y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de  $X$ 's. ¿Cual/es diseño/s ser an mejores?
5. (1.5 puntos) En una planta de fabricacion de semiconductores se decide implantar un sistema de control de calidad, a partir de mediciones de la densidad de un compuesto basico. Se muestra a continuacion las medias y rangos para 15 muestras, cada una basada en 6 observaciones.
- Medias: 6226; 6234; 6231; 6233; 6234; 6232; 6233; 6236; 6232; 6231; 6232; 6232; 6235; 6231; 6234
- Rangos: 4.5; 5.6; 3.7; 5.4; 6.8; 6.4; 3.5; 5.6; 3.8; 4.3; 4.6; 5.8; 4.7; 3.7; 4.1
- a) Utilizando todos los datos, calcula los límites de control para las gráficas de  $\bar{X}$  y  $R$ .
  - b) ¿Puede concluirse que el proceso esta bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - c) Tras realizar b), estima la media y desviación típica del proceso.
  - d) Si la densidad media bajase en 2, ¿cual es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cual sería el número medio de muestras a representar hasta detectar el cambio? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 6232.5 y 2, respectivamente).

Examen 20/1/2010. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG. PRESENCIAL.

Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR. MODELO A.  
INDICA tus cálculos, emplea más de 4 decimales, y haz la letra clara CON BOLÍGRAFO.

TODAS las respuestas deben RAZONARSE.

1. (1.5 pts) Queremos comparar el número de entradas en cierta web (que vamos a suponer Normal) en 2 años. Disponemos del número de entradas diarias del mes de febrero de 2008 (que fue bisiesto), y el del mes de febrero de 2009. El programa que contabiliza las entradas sólo nos ha suministrado la media y desviación típica (ésta aparece dentro del paréntesis) de las entradas diarias en cada uno de esos febreros. En 2008: 30 (3.2), mientras que en 2009: 34 (4.1).
  - a) Mediante el intervalo de confianza al 95 % adecuado, determina si hay cambios en la variación de entradas.
  - b) Mediante el intervalo de confianza al 95 % adecuado, determina si hay cambios en el número medio de entradas.
  
2. (1 pt.) Se ha cargado información en un aparato de almacenamiento de memoria de disco único con diez pistas concéntricas, de tal forma que la probabilidad que el brazo de acceso se requiera a la siguiente pista de acceso  $i$  está dada por  $p_i = (5'5 - |i - 5'5|)/30$ , con  $i = 1, \dots, 10$ , es decir, las  $p_i$  son 1/30, 2/30, 3/30, 4/30, 5/30, 5/30, 4/30, 3/30, 2/30 y 1/30. Una muestra de 200 números de pista aparecen en la tabla siguiente. Determina si la información es consistente con las probabilidades de acceso anteriores (usa  $\alpha = 0.1$ ).

Número de pista	1	2	3	4	5	6	7	8	9	10
Observados	4	15	23	25	38	31	32	14	10	8

3. (1.25 pts; 1.75 pts sin los problemas intentados) Un banco se pregunta si la eliminación de la cuota anual de la tarjeta de crédito de los clientes que carguen en la tarjeta un mínimo de 6000 euros al año, haría aumentar los cargos en sus tarjetas (que asumiremos normales). El banco hace esa oferta a una muestra aleatoria de 200 clientes con tarjeta de crédito. Posteriormente, el banco compara lo que han cargado estos clientes este año con lo que cargaron el año anterior. El resumen descriptivo de la muestra es el siguiente (la variable 'Año actual - Año anterior' es la resta de ambas):

Variable	N	Media	Mediana	Desviación típica
Año anterior	200	12280	12100	3048
Año actual	200	12810	12690	3119
Año actual - Año anterior	200	530	470	140

- a) ¿Existe evidencia significativa a un nivel de  $\alpha = 0.01$  de que la cantidad media cargada aumenta con la oferta de eliminación de la cuota? Indica razonadamente también cuánto piensas que podría valer el p-valor en este caso.
  - b) Si queremos estimar el cargo medio con la oferta, con un error inferior a los 300 euros con una confianza del 95 %, y teniendo en cuenta que podríamos estimar la desviación típica con los datos de la tabla, ¿cuántos clientes deberían muestrearse?
4. (1.5 puntos) En el artículo "A usability comparison of three alternative message formats for an SMS banking service" de Int. J. Human-Computer Studies (2008), se estudió la usabilidad (*usability*) de tres tipos de formatos de mensajes de entrada: Abreviaturas, Números y Formato libre en un servicio de banca por SMS. Vamos a simplificarlo y considerar únicamente los datos de 5 sujetos que llevaron a cabo tres tareas para los 3 formatos aleatoriamente. Los tiempos totales (que asumiremos normales) que les costó realizar las tareas en cada formato de SMS son los siguientes:

Formato	Sujetos				
	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5
Abreviaturas	112	121	259	200	150
Números	51	140	169	208	133
Formato libre	234	161	623	432	406

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre los formatos.

#### Análisis de Varianza

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Formato			79403,4	
Sujetos				
RESIDUAL	56401,9			
TOTAL	318534			

- i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- ii) Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Cuál/es formato/s serían mejores?
5. (1.75 puntos) Los paquetes de galletas de fibra de *Monstruos S.A.* vienen marcados con un contenido de peso neto de 500gr. En el proceso de fabricación se seleccionan 5 paquetes cada día durante dos semanas. Las descriptivas de los pesos (en gramos) obtenidos se recogen a continuación.
- Medias: 503; 497; 513; 499; 497; 505; 503; 504; 498; 500; 505; 499; 502; 494
- Rangos: 13; 25; 26; 4; 34; 17; 6; 12; 20; 15; 7; 10; 24; 9
- Indica razonadamente qué gráficas son las apropiadas para comprobar si el proceso está bajo control, y utilizando todos los datos, calcula los límites de control para dichas gráficas.
  - ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - Tras realizar b), estima la media y desviación típica del proceso.
  - Si el contenido medio bajase en 3 grs., ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el cambio? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 501 y 6, respectivamente).
  - Si ahora se comprueban diariamente 50 paquetes para controlar la rotura de las galletas, descartándose los paquetes con alguna galleta rota, ¿qué tipo de gráficas usarías para controlar la calidad en este caso?

Examen 15/9/2010. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG. PRESENCIAL.  
 Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR.  
 INDICA tus cálculos, emplea más de 4 decimales, y haz la letra clara CON BOLÍGRAFO.

TODAS las respuestas deben RAZONARSE.

1. (1 pt.) En 2008 se llevó a cabo un estudio a adolescentes chilenos titulado “*Usuarios habituales de videojuegos*”. Uno de los aspectos que se estudió entre los que jugaban regularmente, fue la frecuencia o cantidad de tiempo dedicado a jugar según el sexo, cuyos datos se recopilan seguidamente. Determina si son independientes estas dos variables, con  $\alpha = 0.05$ .

Sexo	Frecuencia de juego		
	Diariamente	Habituales (al menos 3 días a la semana)	Sólo el fin de semana
Mujer	3	39	60
Hombre	24	212	111

2. (1.5 puntos) Uno de los atractivos naturales más populares del Parque Nacional Yellowstone en EEUU, es el géiser Old Faithful, que significa “Viejo fiel”, por lo predecible de los intervalos entre sus erupciones. Tenemos en la tabla siguiente intervalos de tiempo (en minutos y que asumiremos normales) entre erupciones para cuatro años distintos (basado en datos del geólogo Rick Hutchinson y el Servicio Nacional de Parques de EEUU):

Año	Observaciones				
1951	60	62	66	52	74
1985	88	57	79	58	85
1995	85	86	79	94	62
1996	94	89	88	91	68

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere saber si el viejo fiel está cambiando, si hay diferencias significativas entre los años.

**Análisis de Varianza**

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Año			513,25	
RESIDUAL				
TOTAL	3686,55			

- i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha=0.05$ ), explicando el resultado que obtengas.
- ii) Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿En qué año/s tenían que esperar menos (en media) los turistas para ver erupciones?
3. (1.5 pts) La entrada en vigor de la Ley 28/2005 por la que se prohíbe fumar en lugares públicos suscitó opiniones muy encontradas. Con la aplicación de la ley, un porcentaje de fumadores decidió dejar el tabaco. En el mes siguiente a la entrada en vigor de la ley, se realizó una encuesta a una muestra aleatoria de 150 fumadores, de los que 33 manifestaron su intención de dejar de fumar.

- a) Calcula un intervalo de confianza del 95 % para la proporción de fumadores que, en toda la población, quisieron dejar de fumar a raíz de la ley.
- b) Calcula el tamaño muestral que se necesita si queremos estimar la proporción de fumadores que quería dejar de fumar con una precisión de  $\pm 0.02$  con un nivel de confianza del 95 %. Utiliza la información de que dispones del enunciado.
- c) El CIS realizó encuestas antes y después de la implantación de la ley que permiten comparar la incidencia de tabaquismo. Cada encuesta se basa en muestras aleatorias de 1500 personas, siendo ambas muestras independientes. En la primera encuesta, realizada antes de la implantación de la ley, el 25.8 % de los encuestados se declaraba fumador. En la segunda encuesta, realizada un año después de implantada la ley, el 23,7 % se declaró fumador. ¿Hay evidencia suficiente para afirmar que ha disminuido la incidencia de tabaquismo tras la entrada en vigor de la ley? (Usa  $\alpha = 0.1$ ).
4. (1 punto) Se quieren comparar dos algoritmos (A y B) para contar el número de apariciones de una cadena de texto. Para ello se seleccionaron al azar 121 ficheros de texto que se ejecutaron para ambos algoritmos, midiéndose el tiempo CPU que podemos asumir Normal. El resumen descriptivo de la muestra es el siguiente (la variable 'A - B' es la resta de ambas):

Variable	N	Media	Mediana	Desviación típica
A	121	235.345	234	20.29
B	121	222.782	223	19.737
A - B	121	12.563	11	12.452

- a) ¿Hay diferencia significativa a un nivel de  $\alpha = 0.05$  entre los tiempos medios de los dos algoritmos? Utiliza el intervalo de confianza apropiado.
- b) Calcula el intervalo de confianza al 99 % para el tiempo medio de ejecución del algoritmo B.
5. (2 pts; 2.5 pts sin los problemas intentados) Pizza Loca ofrece entregas a domicilio y anuncia que el pedido será cortesía de la casa si no se entrega en menos de 30 minutos. Para vigilar la calidad de la entrega, la pizzería registra los tiempos de entrega de seis pedidos seleccionados aleatoriamente cada día durante 2 semanas. Las descriptivas de los tiempos de entrega (en minutos) se listan a continuación.
- Medias: 19.17; 21.33; 23.67; 18.5; 24.5; 18; 17.33; 18.5; 19.83; 15.67; 22; 19.67; 27.83; 22.5
- Rangos: 14; 15; 17; 15; 20; 19; 10; 16; 11; 9; 17; 14; 9; 11
- a) Indica razonadamente qué gráficas son las apropiadas para comprobar si el proceso está bajo control, y utilizando todos los datos, calcula los límites de control para dichas gráficas.
- b) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
- c) Tras realizar b), estima la media y desviación típica del proceso.
- d) Si en representaciones de días posteriores detectáramos una tendencia descendente, ¿debería modificarse el proceso para hacer que las medias fueran estables?
- e) ¿Qué porcentaje de veces el pedido será cortesía de la casa?
- f) Si el tiempo medio de entrega subiese en 3 minutos, ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el cambio? (Si no hubieras realizado el apartado c), asume que la media y desviación típica valdrían 21 y 5, respectivamente).

Examen 19/1/2011. IG23 AMPLIACIÓN DE ESTADÍSTICA. ITIG. PRESENCIAL.  
 Escribe el nombre en todas las hojas incluido el enunciado QUE HAS DE ENTREGAR.  
 INDICA tus cálculos, emplea más de 4 decimales, y haz la letra clara CON BOLÍGRAFO.

TODAS las respuestas deben RAZONARSE.

1. (1 pt.) La compañía fabricante de los caramelos *M&M* asegura que los colores de sus dulces están distribuidos según los porcentajes: 30 % marrón, 20 % amarillo, 20 % rojo, 10 % naranja, 10 % verde y 10 % azul. Se escoge una muestra aleatoria de 100 caramelos, observándose las frecuencias de la tabla siguiente. Verifica si estos datos se ajustan a la distribución de colores que la compañía asegura (usa  $\alpha = 0.05$ ).

Color	Marrón	Amarillo	Rojo	Naranja	Verde	Azul
Observados	33	26	21	8	7	5

2. (1.5 puntos) La empresa encargada de la recolección de residuos urbanos en una ciudad de Arizona, realizó un estudio en el comienzo de la separación de residuos por categoría, para planificar qué equipo necesitaría para recolectar y procesar los residuos. Recolectó durante una semana en hogares seleccionados aleatoriamente (nosotros sólo consideraremos 5), los pesos (asumidos normales) en libras de cuatro categorías de residuos: metal, papel, plástico y vidrio.

Tipo de residuos	Hogares				
	Hogar 1	Hogar 2	Hogar 3	Hogar 4	Hogar 5
Metal	1.09	1.04	2.57	3.02	1.5
Papel	2.41	7.57	9.55	8.82	8.72
Plástico	0.27	1.41	2.19	2.83	2.19
Vidrio	0.83	3.46	4.52	4.92	6.31

A partir de estos datos se ha construido la siguiente tabla ANOVA, pues se quiere estudiar si hay diferencias significativas entre los pesos medios de los tipos de residuos.

Análisis de Varianza

Fuente	Suma Cuadrados	GL	Cuadrados medios	F
Residuos				
Hogares			10,0023	
RESIDUAL	17,1924			
TOTAL	162,264			

- i) Completa los huecos de la tabla ANOVA anterior. Especifica el modelo y realiza la prueba (usa  $\alpha = 0.05$ ), explicando el resultado que obtengas.
- ii) Calcula la LSD al 95 % y realiza las comparaciones explicando los resultados. Identifica grupos homogéneos con columnas de X's. ¿Hay algún/os tipos de residuos que constituya/n una porción especialmente grande del problema del manejo de residuos?
3. (2 pts) Recientemente ha entrado en vigor la nueva ley Antitabaco. Las siguientes son mediciones (que asumiremos normales) de la Encuesta Nacional de Examen de Salud y Nutrición de EEUU, en concreto son niveles de cotinina en el suero (en ng/ml), un metabolito de la nicotina (cuando el cuerpo absorbe nicotina,

se produce cotinina), tanto para no fumadores (no expuestos a humo de tabaco ni en su hogar ni en su trabajo), como para fumadores pasivos (no fumadores pero que están expuestos a humo de tabaco ambiental en su hogar o en su trabajo).

No fumadores	0.07	0.08	0.15	0.24	0.49	0.03	0.05
Fumadores pasivos	0.12	0.38	1.37	3.76	10.23	4.78	2.74

- a) Comprueba razonadamente si las varianzas de los niveles de cotinina difieren, calculando el intervalo de confianza al 90 % adecuado.
  - b) ¿Hay diferencias significativas entre los niveles medios de cotinina para no fumadores y fumadores pasivos? Emplea el intervalo de confianza al 90 % adecuado.
  - c) Centrándonos únicamente en los fumadores pasivos, si deseamos que el error en la estimación del nivel de cotinina medio sea inferior a 0.2 con una confianza del 90 %, y usando la muestra anterior para estimar  $\sigma$ , ¿cuántos sujetos fumadores pasivos deberían medirse como mínimo?
4. (1 punto) En 1954, en las pruebas iniciales de la vacuna Salk, se realizó un experimento para probar la efectividad de la vacuna Salk como protección contra los devastadores efectos de la poliomielitis. A 200.000 niños se les inyectó un placebo (una solución salina inocua), de los cuales 115 contrajeron posteriormente polio. A otros 200.000 niños se les vacunó, de éstos 33 contrajeron la enfermedad. El experimento fue doble ciego (ni los niños ni los doctores sabían si se les suministraba la vacuna o el placebo). Con un nivel de significación de 0.01, prueba la afirmación de que la vacuna Salk es efectiva para disminuir la proporción de casos de polio. ¿Es efectiva la vacuna?
5. (2 pts; 2.5 pts sin los problemas intentados) La casa de la moneda de EEUU tiene como objetivo acuñar monedas de 25 centavos de dólar con un peso de 5.67 g., aunque cualquier peso entre 5.443 y 5.897 g. se considera aceptable. Una máquina nueva de acuñar entra en servicio y se registran los pesos de 5 monedas seleccionadas aleatoriamente cada 15 minutos durante 4 horas consecutivas. Las descriptivas de los pesos (en gramos) se listan a continuación.
- Medias: 5.66; 5.65; 5.7; 5.64; 5.69; 5.63; 5.61; 5.62; 5.73; 5.66; 5.59; 5.72; 5.66; 5.76; 5.73; 5.66
- Rangos: 0.68; 0.1; 0.08; 0.31; 0.07; 0.16; 0.1; 0.14; 0.29; 0.13; 0.17; 0.72; 0.24; 0.58; 0.12; 0.05
- a) Indica razonadamente qué gráficas son las apropiadas para comprobar si el proceso está bajo control, y utilizando todos los datos, calcula los límites de control para dichas gráficas.
  - b) ¿Puede concluirse que el proceso está bajo control? De no ser así, supón que pueden encontrarse las causas asignables, y recalcula los límites de control una vez eliminados los puntos fuera de control.
  - c) Tras realizar b), estima la media y desviación típica del proceso.
  - d) ¿Qué porcentaje de monedas no será aceptable?
  - e) Calcula el  $ICP$  y el  $ICP_k$ , e interprétalos.
  - f) Si hubiera un desajuste y el peso medio bajara en 0.07 gramos, ¿cuál es la probabilidad de detectar el cambio en la muestra siguiente? ¿Cuál sería el número medio de muestras a representar hasta detectar el cambio?
6. (1 punto) RECUPERACIÓN DEL PUNTO DEL PROYECTO. SE ELEGIRÁ LA NOTA MÁS ALTA ENTRE EL PROYECTO Y EL SIGUIENTE PROBLEMA. POR TANTO, SI TIENES UN PUNTO POR EL PROYECTO, NO NECESITAS REALIZARLO.

Escribe el código completo en R que deberías emplear para resolver los problemas 1 y 2.

7. (1 punto) RECUPERACIÓN DEL PUNTO POR LAS MEMORIAS DE PRÁCTICAS. SE ELEGIRÁ LA NOTA MÁS ALTA ENTRE LAS MEMORIAS Y EL SIGUIENTE PROBLEMA. POR TANTO, SI TIENES UN PUNTO POR LAS MEMORIAS, NO NECESITAS REALIZARLO.

Escribe el código completo en R que deberías emplear para resolver los problemas 3 (a) y b)) y 4.

**Problema 1. (2'25 puntos)** Se toma una muestra de estudiantes universitarios de informática de dos universidades y se les pregunta por su navegador de internet preferido. En la universidad A, 70 prefieren el “Exploder” de 200 encuestados; y en la universidad B, 93 prefieren el “Exploder” de 300 encuestados.

**1.a)** Calcula el intervalo de confianza del 95% para la diferencia de proporciones de preferencia de ambas universidades e **interpretálo**.

**1.b)** Consideramos ahora sólo la universidad A. Queremos estimar la proporción de “preferencia del Exploder” con un error de cómo mucho 0'03 al 95% de confianza. Calcula el tamaño muestral necesario, utilizando la proporción anterior como muestra preliminar.

**1.c)** Los 200 encuestados de la universidad A pertenecían a tres carreras distintas: II, ITIS e ITIG. En la encuesta tenían dos opciones de navegador a elegir: “Exploder” y “Notescapes”, de forma que los datos desglosados aparecen en la tabla siguiente:

Navegador	II	ITIS	ITIG
<i>Exploder</i>	15	30	25
<i>Notescapes</i>	25	50	55

Realiza un contraste de hipótesis al nivel de significación del 5% para comprobar la independencia entre la carrera y la preferencia por un navegador u otro.

**Problema 2. (1'25 puntos)** Para comparar los niveles de ruido de fondo en un entorno de producción, se toman 15 medidas aleatorias (independientes) en dos lugares de la planta, que asumiremos normales. Los datos dieron:

- para la posición A → media de 82.4 y varianza de 2.3
- para la posición B → media de 78.3 y varianza de 4.1

Si las condiciones en los dos sitios difieren, especialmente con respecto al tipo de maquinaria en operación en cada lugar, podría ser arriesgado suponer varianzas iguales, por tanto asumiremos varianzas distintas.

Queremos contrastar, al nivel de significación del 5%, que el nivel de ruido medio en la posición A es 3 decibelios mayor que el nivel de ruido medio en la posición B.

**2.a)** Identifica la información muestral que tienes.

**2.b)** Plantea las hipótesis nula y alternativa adecuadas.

**2.c)** Resuelve el contraste de hipótesis que has planteado y toma la decisión correspondiente.

**Problema 3. (1'75 puntos)** Se controla el grosor de circuitos electrónicos (en pulgadas) utilizando los gráficos X-bar y R. Tomamos 20 muestras de tamaño 5 y obtenemos los siguientes datos:

$$\sum x_i = 1'2 \qquad \sum r_i = 0'27$$

**3.a)** Determina cuánto valen **m** y **n**. Encuentra los límites de control de los gráficos X-bar y R

**3.b)** Suponiendo que el proceso está bajo control, estima la media y desviación típica del proceso.

**3.c)** Supongamos que el grosor tiene que estar entre 0'055 y 0'076.

**3.c.1)** ¿Cuál es el porcentaje de unidades defectuosas que se está produciendo en este proceso?

**3.c.2)** Calcula el ICP y el  $ICP_k$  e interprétalos.

**3.d)** Supongamos que la media se corre a 0'065.

**3.d.1)** ¿Cuál es la probabilidad de que el desplazamiento se detecte en la primera muestra después del corrimiento?

**3.d.2)** ¿Cuál es la ARL después del corrimiento?

**Problema 4. (1'75 puntos)** Se pretende comparar tres programas de traducción automática (A, B y C) respecto a la ocupación de memoria de sus diccionarios. Para tres muestras aleatorias independientes de textos de igual tamaño ha hecho la traducción empleando el programa A, B y C sobre un PC aislado y se ha medido el espacio de memoria (en miles de bytes) consumidos en cada ejecución, obteniendo los siguientes resultados:

Diccionario				
A	1803	1850	1700	1940
B	2105	2010	2120	2180
C	2290	2098	2240	2310

A partir de estos datos se ha construido la siguiente tabla ANOVA, para estudiar si hay diferencias significativas en la ocupación media de memoria de cada programa.

Fuente variación	Suma cuadrados	Grados libertad	Media cuadrática	Estadístico F
Factor		$\alpha - 1 =$	$MC_{tr} =$	$F =$
Error		$N - \alpha =$	$MCE = 8028'72$	
<b>TOTAL</b>	<b>425462</b>	<b><math>N - 1 =</math></b>		

**4.a)** Identifica el factor con el que trabajas, el número de niveles  **$\alpha$**  del factor, el número de datos por nivel y el número de datos total **N**.

**4.b)** Completa los huecos de la tabla ANOVA anterior. Especifica el modelo ANOVA que aplicas y realiza la prueba (usa  $\alpha = 0'05$ ) explicando el resultado que obtengas.

**4.c)** Calculas las medias de los diferentes niveles y aplica la LSD al 95% para realizar las comparaciones. **Explica los resultados.**