Apellidos: **Computadores** ☐

Nombre: **Software** ☐

Estadística. 3º examen parcial. 20-01-2014

Test (20 % de la nota del examen)

- Tiempo para esta parte del examen: 1 hora y 30 minutos.
- El test se recogerá a los 40 minutos.
- En cada pregunta de test, una y sólo una de las respuestas (a), (b) y (c) es cierta. Poner la letra elegida o dejar en blanco.
- Calificación: acierto = +1, fallo = −1/2 y blanco = 0.

Sea $X \sim B(5, p)$, con p desconocido. Sea $[X_1, X_2]$ una muestra de X. Se toma $\widehat{p} = 3\,X_1 - 2\,X_2$ como estimador de p. El error cuadrático medio de $\widehat{p}$ vale:

(a) $16\,p^2 + 65\,p\,(1-p)$*

(b) $4\,p^2 + 25\,p\,(1-p)$

(c) $5\,p\,(1-p)$

Solución

La media de $\widehat{p}$ vale:

$$\mu\,(3\,X_1 - 2\,X_2) = 3\,\mu\,(X_1) - 2\,\mu\,(X_2) = \mu\,(X) = 5\,p.$$

Por tano, $\widehat{p}$ no es centrado. La varianza de $\widehat{p}$ vale:

$$\sigma^2\,(3\,X_1 - 2\,X_2) = 3^2\,\sigma^2\,(X_1) + (-2)^2\,\sigma^2\,(X_2) = 13\,\sigma^2\,(X) = 13 \cdot 5\,p\,(1-p)\,,$$

El error cuadrático medio de $\widehat{p}$ es el cuadrado del sesgo más la varianza, que vale:

$$\mathrm{ECM}\,(\widehat{p}) = (\mu\,(\widehat{p}) - p)^2 + \sigma^2\,(\widehat{p}) = (5\,p - p)^2 + 65\,p\,(1-p) = 16\,p^2 + 65\,p\,(1-p)\,.$$

Sea $X \sim N(\mu, \sigma)$, con σ desconocida. Se toma una muestra de X con $\overline{x} = 2.3$ y $s^2 = 1.5$. Se construye un intervalo de confianza para σ^2 y se obtiene:

$$\left(\frac{(n-1)\,s^2}{b}, \frac{(n-1)\,s^2}{a}\right) = (1.2, 3.4)\,.$$

Entonces, la cota de error más ajustada que se puede dar al estimar σ^2 por s^2 vale:

(a) 0.3

(b) 1.1

(c) 1.9*

Solución

El error de la estimación está acotado por la distancia de $s^2 = 1.5$ al extremo más lejano del intervalo de confianza. En este caso, vale $3.4 - 1.5 = 1.9$.

Se toma una muestra de $X \sim \text{Exp}(\beta)$ con tamaño $n = 300$. ¿Se puede construir un intervalo de confianza aproximado para la desviación típica de X?

(a) No, porque X no es normal.

(b) Sí, usando el pivote aproximado $(n-1) S^2/\sigma^2 \approx \chi^2_{n-1}$, ya que n es grande.

(c) Sí, usando el pivote aproximado $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \approx N(0,1)$, ya que n es grande y la desviación típica coincide con la media.*

Solución
En general, no se puede construir un intervalo de confianza para la varianza de una variable no normal. Pero en este caso, como n es grande, se puede aproximar $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \approx N(0,1)$ y usarlo como pivote para construir un intervalo de confianza aproximado para $\mu = 1/\beta = \sigma$. La aproximación $(n-1) S^2/\sigma^2 \approx \chi^2_{n-1}$ no está garantizada aunque n sea grande.

•

Se realiza un contraste de hipótesis. Si al 4 % de significación se acepta H_0, entonces al 5 % de significación:

(a) Se rechaza H_0.

(b) Se rechaza o acepta H_0 dependiendo del p-valor del contraste.*

(c) Se acepta H_0.

Solución
Aceptar H_0, al 4 % de significación, significa que p-valor > 0.04. Pero $0.04 < 0.05$. No se puede decidir, al 5 % de significación, sin más información sobre el p-valor.

•

Se analiza el número de veces que cierto componente está funcionando en 4 observaciones del mismo. Se observa una muestra de tamaño 80 y se quiere estudiar si los datos proceden de una distribución $B(4,p)$, al 5 % de significación. Se realiza el contraste adecuado y resulta:

Nº veces	Frecuencia	p_i	$n \cdot p_i$	$n_i^2/(n \cdot p_i)$	$(n_i - n \cdot p_i)^2/(n \cdot p_i)$
0	13	0.2401	19.208	8.7984	2.0064
1	41	0.4116	32.928	51.0508	1.9787
2	23	0.2646	21.168	24.9906	0.1585
≥ 3	3	0.0837	6.696	1.3441	2.0400
	$n = 80$			86.1838	

El p-valor del contraste verifica:

(a) p-valor > 0.025 y se acepta que los datos vengan de una distribución $B(4,p)$.

(b) p-valor < 0.05 y se rechaza que los datos vengan de una distribución $B(4,p)$.*

(c) p-valor > 0.05 y se acepta que los datos vengan de una distribución $B(4,p)$.

Solución
El valor del pivote es $d = 86.1838 - 80 = 6.1838$ (o bien $2.0064 + 1.9787 + 0.1585 + 2.0400 = 6.1836$).

Por tanto, el p-valor del contraste es:

$$p\text{-valor} \simeq p\left(\chi^2_{4-1-1} > 6.1838\right) = 1 - p\left(\chi^2_2 \leq 6.1838\right) \begin{cases} < 1 - 0.95 = 0.05 \\ > 1 - 0.975 = 0.025 \end{cases}$$

En consecuencia, la decisión es: p-valor $< 0.05 = \alpha \Rightarrow$ rechazar H_0. Al $5\,\%$ de significación, se rechaza que los datos vengan de una distribución $B\,(4, p)$.

Estadística. 3º examen parcial. 20-01-2014

Instrucciones:

- Tiempo para esta parte del examen: 1 hora y 30 minutos.
- Entregar la teoría y cada problema en hojas separadas.
- Sólo se puede salir al servicio en casos excepcionales y previa autorización de un profesor.
- Las soluciones, las notas y la fecha de revisión del examen se publicarán en el espacio Moodle de la asignatura.

Teoría ($10\,\%$ de la nota del examen)

Sea $[X_1, \ldots, X_n]$ una muestra aleatoria simple de la variable aleatoria X. Demostrar que:

- La esperanza de $\overline{X}$ coincide con la esperanza de X.
- La varianza de $\overline{X}$ coincide con la varianza de X dividida por el tamaño muestral.

Problema 1 ($35\,\%$ de la nota del examen)

El tiempo que transcurre desde que cierto modelo de servidor se estropea hasta que es reparado, en horas, es una variable aleatoria con densidad:

$$f(x) = \begin{cases} 0 & \text{si } x < 3 \\ \alpha\, 3^\alpha / x^{\alpha+1} & \text{si } x \geq 3 \end{cases} \qquad \alpha > 2.$$

Se toma una muestra aleatoria simple de dicho tiempo, con tamaño $n = 256$, media $4.6\,\mathrm{h}$, cuasivarianza $4\,\mathrm{h}^2$ y $\sum_{i=1}^{n} \ln(x_i) = 373.396$. Se pide:

(a) (1.5 puntos) Obtener el estimador y la estimación de α por el método de los momentos.

(b) (2 puntos) Obtener el estimador y la estimación de α por máxima verosimilitud.

(c) (3 puntos) Construir razonadamente, a partir del pivote correspondiente, un intervalo, al $90\,\%$ de confianza, para el tiempo medio que transcurre hasta que el servidor es reparado. Estimar el tiempo medio que transcurre hasta que el servidor es reparado. Acotar el error absoluto de la estimación.

(d) (1 punto) Manteniendo el nivel de significación y suponiendo que la cuasivarianza muestral permanece constante, calcular el tamaño de la muestra necesario para conseguir que el error absoluto de la estimación anterior sea menor que 0.1 horas.

(e) (2.5 puntos) En este tipo de servidor, el tiempo de ejecución de un proceso, en segundos, sigue una distribución normal con desviación típica $0.2\,\mathrm{s}$. Se han hecho ciertos cambios en el servidor y se sospecha que la variabilidad en el tiempo de ejecución ha aumentado. Se toma una muestra de tamaño 20, resultando que la cuasivarianza muestral es $0.07\,\mathrm{s}^2$. A la vista de los datos y con un nivel de significación del $5\,\%$ ¿es cierta la sospecha que se tiene? Obtener el p-valor.

Solución

Apartado (a)

Se considera la variable aleatoria:

$$X : \text{"tiempo desde que el servidor se estropea hasta que es reparado (h)"} \sim \text{Pareto}(\alpha, 3).$$

La función de densidad de X corresponde a una distribución de Pareto con parámetros α y $k = 3$.

Si $\alpha > 2$, su estimador por momentos se despeja de la ecuación:

$$\mu(X) = \frac{\alpha\, k}{\alpha - 1} = \overline{X} \Rightarrow \alpha\, k = \alpha\, \overline{X} - \overline{X} \Rightarrow \overline{X} = \alpha\, \overline{X} - \alpha\, k = \alpha\left(\overline{X} - k\right) \Rightarrow \widehat{\alpha} = \frac{\overline{X}}{\overline{X} - k}.$$

Con los datos muestrales, la estimación de α por momentos es $\widehat{\alpha} = \dfrac{4.6}{4.6 - 3} \simeq 2.875$.

Apartado (b)

Observada la muestra $[x_1, \ldots, x_n]$, la función de verosimilitud y su logaritmo valen:

$$L(\alpha) = \left(\prod_{i=1}^{n} \frac{\alpha\, k^{\alpha}}{{x_i}^{\alpha+1}}\right) \Rightarrow \ln(L(\alpha)) = \sum_{i=1}^{n} \left(\ln(\alpha) + \alpha \ln(k) - (\alpha+1)\ln(x_i)\right).$$

La derivada de esta expresión respecto a α vale:

$$\frac{d}{d\alpha} \ln(L(\alpha)) = \sum_{i=1}^{n} \left(\frac{1}{\alpha} + \ln(k) - \ln(x_i)\right) = \frac{n}{\alpha} + n\, \ln(k) - \sum_{i=1}^{n} \ln(x_i).$$

Igualando a 0 y despejando α, resulta:

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \ln(x_i) - n\, \ln(k) \Rightarrow \widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(x_i) - n\, \ln(k)}.$$

Con los datos muestrales, la estimación de α por máxima verosimilitud es:

$$\widehat{\alpha}_2 = \frac{256}{373.396 - 256\, \ln(3)} \simeq 2.77804.$$

SG ajusta $X \sim \text{Pareto}\,(2.77803, 3)$ con p-valor $= 0.960253$, luego el ajuste es bueno.

Apartado (c)

Se trata de construir un intervalo, al 90 % de confianza, para la media de X:

1. Normalidad: como X sigue una distribución de Pareto, no se puede suponer que sea normal. No obstante, la muestra de X es grande ($n = 256$), luego $\overline{X}$ es aproximadamente normal.
2. Pivote aproximado: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \approx N(0.1)$.
3. Constantes: $p(N(0.1) \le a) = 0.95 \Rightarrow a = 1.65$.
4. Pivotar: $\mu \in 4.6 + 1.65\, \dfrac{\sqrt{4}}{\sqrt{256}}\, [-1, 1] \simeq [4.39375, 4.80625]$.

La media poblacional se estima con la media muestral, que es el centro del intervalo de confianza: $\widehat{\mu} = \overline{x} = 4.6\,\text{h}$. El error absoluto de la estimación está acotado por el radio de dicho intervalo, que vale $1.65\, \dfrac{\sqrt{4}}{\sqrt{256}} \simeq 0.20625$.

Apartado (d)

Se trata de construir un intervalo, al 90 % de confianza, para μ con radio menor que 0.1:

$$a\, \frac{s}{\sqrt{n}} < 0.1 \Rightarrow n > \left(\frac{a \cdot s}{0.1}\right)^2 = \frac{1.65^2 \cdot 4}{0.1^2} \simeq 1089.0 \Rightarrow n \ge 1090.$$

Se supone que s^2 no cambia.

Apartado (e)

Se considera la variable aleatoria:

$$X: \text{"tiempo de ejecución de un proceso en el servidor (s)"} \sim N(\mu, \sigma).$$

Se trata de contrastar, al $5\,\%$ de significación, que la desviación típica de X verifica $\sigma > 0.2$:

$$\begin{cases} H_0 : \sigma \leq 0.2 \\ H_1 : \sigma > 0.2 \end{cases} \Rightarrow \begin{cases} H_0 : \sigma^2 \leq 0.2^2 \\ H_1 : \sigma^2 > 0.2^2 \end{cases} \Rightarrow \begin{cases} H_0 : \sigma^2 = 0.04 \\ H_1 : \sigma^2 > 0.04 \end{cases}$$

1. Normalidad: se supone.
2. Pivote: $\dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1} = \chi^2_{19}$.
3. p-valor $= p\left(S^2 \geq 0.07 \mid \sigma^2 = 0.04\right) = p\left(\dfrac{(n-1)\,S^2}{\sigma^2} \geq \dfrac{19 \cdot 0.07}{0.04}\right) =$

$$= p\left(\chi^2_{19} \geq 33.25\right) = 1 - p\left(\chi^2_{19} \leq 33.25\right) \begin{cases} < 1 - 0.975 = 0.025 \\ > 1 - 0.99 = 0.01 \end{cases} \qquad \text{(SG 0.0224918).}$$

4. Decisión: p-valor $< 0.025 < 0.05 = \alpha \Rightarrow$ rechazar H_0.

Al $5\,\%$ de significación, se acepta que la variabilidad en el tiempo de ejecución ha aumentado.

Apellidos:	
Nombre:	**Titulación:**

Universidad Politécnica de Madrid — ETSI SISTEMAS INFORMÁTICOS	**Estadística. Tercer parcial.** **Parte con ordenador.**	**20-01-2014**
	• **Tiempo: 50 minutos** • **Valor: 35%**	**MODELOS A y B**

Dos aseguradoras A y B, de fuerte implantación, han realizado un estudio sobre los costes de reparación de los vehículos asegurados en cierta ciudad. Los datos se han recogido durante los días laborables de un año (252) y están contenidos en las variables X(A), X(B), F(A) y F(B) del fichero Enero2014.sf3.

La variable X(A) contiene el número de vehículos asegurados por A que llegan diariamente a los talleres para ser reparados.

(a) (2 puntos) Obtener un intervalo de confianza al 92% para el número medio de vehículos asegurados por A que llegan a los talleres para ser reparados. Es necesario detallar todas las hipótesis necesarias para construir el intervalo. ¿Se puede admitir que llegan a los talleres una media de 8 vehículos diarios para reparar?

La variable X(A) es discreta (la muestra toma valores enteros entre 2 y 16). Contrastamos si se puede considerar normal:

Uncensored Data - X(A)

Data variable: X(A)

252 values ranging from 2,0 to 16,0

Fitted Distributions

Normal
mean = 7,84524
standard deviation = 2,8277

Goodness-of-Fit Tests for X(A)

Kolmogorov-Smirnov Test

	Normal
DPLUS	*0,121467*
DMINUS	*0,0658884*
DN	*0,121467*
P-Value	*0,00117907*

A la vista del resultado no podemos considerar que la variable X(A) es normal. Ahora bien, como la muestra es suficientemente grande (252) podemos aplicar el teorema central del límite y obtener un intervalo aproximado para la media de la población X(A). Calculándolo con Statgraphics, obtenemos:

92,0% confidence interval for mean: 7,84524 +/- 0,313115 [7,53212; 8,15835].

Como 8 pertenece al intervalo, con ese nivel de confianza se puede admitir que llegan 8 vehículos a los talleres.

Para el modelo B, usando un nivel de confianza del 90%:

La variable tampoco se puede considerar normal. Razonando análogamente se obtiene:

90,0% confidence interval for mean: 6,94444 +/- 0,263005 [6,68144; 7,20745]

Como 8 no pertenece al intervalo no se puede admitir que llega una media de 8 vehículos diarios.

La aseguradora A sospecha sobre la facturación de los talleres, por lo que hace un estudio sobre el importe diario que paga a dichos talleres. La variable F(A) recoge los valores de estos importes, en euros. No necesariamente todos los talleres de la ciudad son muestreados todos los días. La aseguradora A afirma que el importe medio diario que paga a los talleres de esa ciudad es de 4000€ con una desviación típica menor de 100€, lo cual le hace sospechar la existencia de algún acuerdo entre los talleres. Utilizando un nivel de significación del 4% (En el Modelo B 6%):

(b) (2.5 puntos) Contrastar las afirmaciones de la aseguradora. Es necesario justificar todas las hipótesis necesarias para efectuar el estudio, así como el planteamiento de los contrastes necesarios para realizarlo.

Debemos plantear dos contrastes:

$$\begin{cases} H_0 : \mu = 4000 \\ H_1 : \mu \neq 4000 \end{cases}; \quad \begin{cases} H_0 : \sigma = 100 \\ H_1 : \sigma < 100 \end{cases} \qquad \begin{cases} H_0 : \mu = 3800 \\ H_1 : \mu \neq 3800 \end{cases}; \quad \begin{cases} H_0 : \sigma = 80 \\ H_1 : \sigma < 80 \end{cases}$$

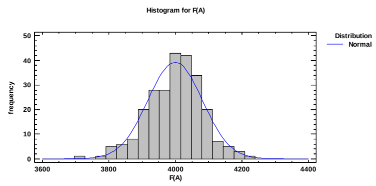
Miramos en primer lugar si la variable F(A) es normal, porque en caso de no serlo, no podríamos hallar el intervalo para la varianza.

Fitted Distributions

Normal
mean = 3999,34
standard deviation = 81,9485

Goodness-of-Fit Tests for F(A)
Kolmogorov-Smirnov Test

	Normal
DPLUS	**0,0317586**
DMINUS	**0,0587412**
DN	**0,0587412**
P-Value	**0,35232**



A la vista de p-valor, asumimos normalidad.
Resolvemos los contrastes con Statgraphics:

<u>t-test</u>
Null hypothesis: mean = 4000,0
Alternative: not equal

Computed t statistic = -0,127658
P-Value = 0,898521
Do not reject the null hypothesis for alpha = 0,04.

<u>chi-square test</u>
Null hypothesis: sigma = 100,0
Alternative: less than

Computed chi-square statistic = 168,561
P-Value = 0,0000168268
Reject the null hypothesis for alpha = 0,04.

Por tanto, con ese nivel de significación, admitimos las sospechas de la aseguradora.

Para el modelo B:

La variable F(B) es normal (P-Value=0,973473). Resolvemos los contrastes:

<u>t-test</u>
Null hypothesis: mean = 3800,0
Alternative: not equal

Computed t statistic = 1,97745
P-Value = 0,0490852
Reject the null hypothesis for alpha = 0,06.

<u>chi-square test</u>
Null hypothesis: sigma = 80,0
Alternative: less than

Computed chi-square statistic = 299,339
P-Value = 0,980391
Do not reject the null hypothesis for alpha = 0,06.

Con ese nivel de significación solo admitimos que la desviación típica es menor que 80€.

(c) (2.5 puntos) Contrastar si hay diferencias significativas en los importes pagados por las aseguradoras A y B a los talleres de reparación de esa ciudad. Es necesario justificar las hipótesis necesarias para efectuar el estudio, así como el planteamiento de los contrastes necesarios para realizarlo.

Nos piden contrastar:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 0 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} \neq 0 \end{cases}$$

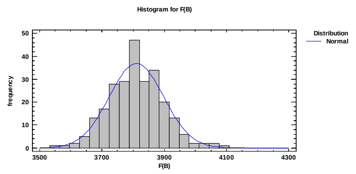
En primer lugar, observamos que las muestras no son pareadas porque, tal y como dice en el enunciado, no todos los talleres reciben coches de ambas aseguradoras. Por tanto estamos en el caso de muestras independientes.

Necesitamos, por tanto, saber si podemos considerar que las varianzas son iguales, para lo cual debemos contrastar si la variable F(B) es normal. Utilizamos Statgraphics:

Fitted Distributions
Normal
mean = 3810,88
standard deviation = 87,3645

Goodness-of-Fit Tests for F(B)
Kolmogorov-Smirnov Test

	Normal
DPLUS	0,0304645
DMINUS	0,0218801
DN	0,0304645
P-Value	0,973473



A la vista del gráfico y el p-valor, admitimos que F(B) es normal.

Contrastamos ahora la igualdad de varianzas:

$$\begin{cases} H_0 : \dfrac{\sigma^2_{F(A)}}{\sigma^2_{F(B)}} = 1 \\ H_1 : \dfrac{\sigma^2_{F(A)}}{\sigma^2_{F(B)}} \neq 1 \end{cases}$$

F-test to Compare Standard Deviations
  Null hypothesis: sigma1 = sigma2
  Alt. hypothesis: sigma1 NE sigma2
  F = 0,879858   P-value = 0,311273
  Do not reject the null hypothesis for alpha = 0,04.

Por tanto, admitimos que las varianzas son iguales.

Contrastamos la igualdad de medias:

t test to compare means
  Null hypothesis: mean1 = mean2
  Alt. hypothesis: mean1 NE mean2
    assuming equal variances: t = 24,9757   P-value = 0
  Reject the null hypothesis for alpha = 0,04.

Por tanto no admitimos, con ese nivel de significación que ambas aseguradoras pagan lo mismo.

Este apartado es igual para el modelo B, con α=0.06.

(d) (1.5 puntos) En caso que haya diferencias significativas entre lo que pagan ambas aseguradoras, contrastar si hay evidencia de que una paga más que la otra.

Contrastamos si la aseguradora A paga más que la B. Como sabemos que podemos considerar las varianzas iguales, contrastamos directamente:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 0 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 0 \end{cases}$$

t test to compare means
  Null hypothesis: mean1 = mean2
  Alt. hypothesis: mean1 NE mean2
    assuming equal variances: t = 24,9757   P-value = 0
  Reject the null hypothesis for alpha = 0,04.

Por tanto, rechazamos H_0 y asumimos que la aseguradora A paga más que la B.

Este apartado es igual para el modelo B, con α=0.06.

(e) (1.5 puntos) Realizar los contrastes necesarios para comprobar (en pasos de 50€) cuánto más paga a los talleres la aseguradora que paga más. En el modelo B los pasos son de 55€.

Vamos a ver si hay evidencia de que la aseguradora A paga más de 50€ más que la B. Planteamos el contraste:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 50 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 50 \end{cases} \qquad \begin{matrix} H_0 : \mu_{F(A)} - \mu_{F(B)} = 55 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 55 \end{matrix}$$

y usamos el mismo nivel de significación:

t test to compare means
  Null hypothesis: mean1 - mean2 = 50,0
  Alt. hypothesis: greater than
    assuming equal variances: t = 18,3494   P-value = 0
  Reject the null hypothesis for alpha = 0,04.

t test to compare means
  Null hypothesis: mean1 - mean2 = 55,0
  Alt. hypothesis: greater than
    assuming equal variances: t = 17,6868   P-value = 0
  Reject the null hypothesis for alpha = 0,06.

Por tanto, A paga más de 50€ más que B. Modelo B: A paga más de 60€ más que B) Repetimos para 100:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 100 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 100 \end{cases}$$

t test to compare means
  Null hypothesis: mean1 - mean2 = 100,0
  Alt. hypothesis: greater than
    assuming equal variances: t = 11,7231   P-value = 0
  Reject the null hypothesis for alpha = 0,04.

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 110 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 110 \end{cases}$$

t test to compare means
Null hypothesis: mean1 - mean2 = 110,0
Alt. hypothesis: greater than
assuming equal variances: t = 10,3978   P-value = 0
Reject the null hypothesis for alpha = 0,06.

Para 150:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 150 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 150 \end{cases}$$

t test to compare means
Null hypothesis: mean1 - mean2 = 150,0
Alt. hypothesis: greater than
assuming equal variances: t = 5,09673   P-value = 2,45125E-7
Reject the null hypothesis for alpha = 0,04.

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 165 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 165 \end{cases}$$

t test to compare means
Null hypothesis: mean1 - mean2 = 165,0
Alt. hypothesis: greater than
assuming equal variances: t = 3,10884   P-value = 0,00099243
Reject the null hypothesis for alpha = 0,06.

Para 200:

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 200 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 200 \end{cases}$$

t test to compare means
Null hypothesis: mean1 - mean2 = 200,0
Alt. hypothesis: greater than
assuming equal variances: t = -1,52959   P-value = 0,936626
Do not reject the null hypothesis for alpha = 0,04.

$$\begin{cases} H_0 : \mu_{F(A)} - \mu_{F(B)} = 220 \\ H_1 : \mu_{F(A)} - \mu_{F(B)} > 220 \end{cases}$$

t test to compare means
Null hypothesis: mean1 - mean2 = 220,0
Alt. hypothesis: greater than
assuming equal variances: t = -4,18012   P-value = 0,999983
Do not reject the null hypothesis for alpha = 0,06.

Ahora no rechazamos H0.
Por tanto, con ese nivel de confianza admitimos que la aseguradora A paga diariamente más de 150€ de media que la B, pero no paga más de 200€ más.
Para el Modelo B admitimos que la aseguradora A paga diariamente más de 165€ de media que la B, pero no más de 220€.

Apellidos: Computadores ☐

Nombre: Software ☐

Estadística. Examen final. 20-01-2014

Test (20 % de la nota del examen)

- Tiempo para esta parte del examen: 3 h.
- El test se recogerá a los 40 minutos.
- En cada pregunta de test, una y sólo una de las respuestas (a), (b) y (c) es cierta. Poner la letra elegida o dejar en blanco.
- Calificación: acierto = +1, fallo = −1/2 y blanco = 0.

Sean A y B sucesos de probabilidad no nula. Si $p(A \mid B) = p(B \mid A)$, entonces:

(a) A y B son incompatibles.

(b) A y B son equiprobables.*

(c) A y B son independientes.

Solución

Por definición de probabilidad condicionada, se verifica:

$$p(A \mid B) = p(B \mid A) \Rightarrow \frac{p(A \cap B)}{p(B)} = \frac{p(B \cap A)}{p(A)} \Rightarrow p(A) = p(B).$$

Por tanto, A y B son equiprobables. Las otras dos afirmaciones son falsas, en general. Un contraejemplo es:

$$E = \{1, 2, 3\} \text{ espacio muestral equiprobable}, \quad A = \{1, 2\}, \quad B = \{2, 3\},$$

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p\{2\}}{p(B)} = \frac{1/3}{2/3} = \frac{1}{2},$$

$$p(B \mid A) = \frac{p(B \cap A)}{p(A)} = \frac{p\{2\}}{p(A)} = \frac{1/3}{2/3} = \frac{1}{2},$$

$$A \cap B = \{2\} \neq \emptyset \Rightarrow A \text{ y } B \text{ son compatibles},$$

$$p(A \mid B) = \frac{1}{2} \neq p(A) = \frac{2}{3} \Rightarrow A \text{ y } B \text{ son dependientes}.$$

———— • ————

Se lanzan 4 dados. La probabilidad de obtener exactamente una pareja (dos dados con el mismo número y los otros dos con números diferentes de dicho número y entre sí) vale:

(a) $\dfrac{\binom{4}{2}}{6^4}$

(b) $\dfrac{1}{6} \cdot \dfrac{5}{6} \cdot \dfrac{4}{6}$

(c) $\binom{4}{2} \dfrac{6 \cdot 5 \cdot 4}{6^4}$*

Solución

Se considera el suceso:

$$A: \text{“obtener exactamente una pareja”} = \{xxyz\}.$$

Su probabilidad vale:

$$p(A) = \binom{4}{2} \frac{6}{6} \frac{1}{6} \frac{5}{6} \frac{4}{6} = \binom{4}{2} \frac{6 \cdot 5 \cdot 4}{6^4}.$$

Tres aulas tienen 3, 4 y 6 ordenadores, respectivamente. Un usuario elige un aula al azar y, dentro de ella, un ordenador al azar. Si en cada aula hay un ordenador con virus, entonces la probabilidad de que el usuario lo elija vale:

(a) 1/13
(b) 1/4*
(c) 1/3

Solución
Se consideran los sucesos:

A_i : "elegir el aula i", $\quad p(A_i) = 1/3, \quad i = 1, 2, 3,$
B : "elejir un ordenador con virus", $\quad p(B \mid A_1) = 1/3, \quad p(B \mid A_2) = 1/4, \quad p(B \mid A_3) = 1/6.$

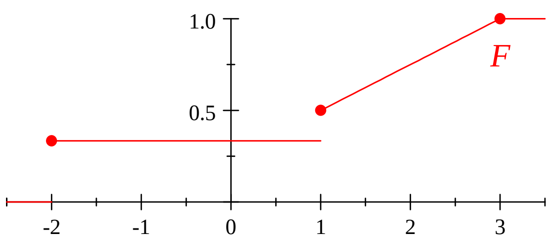
Se trata de calcular la probabilidad de B. Usando la regla de la probabilidad total, resulta:

$$p(B) = \sum_{i=1}^{3} p(A_i)\, p(B \mid A_i) = \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{6} = \frac{1}{4}.$$

La función $F(x) = \begin{cases} 0 & \text{si } x < -2 \\ 1/3 & \text{si } -2 \le x < 1 \\ (x+1)/4 & \text{si } 1 \le x < 3 \\ 1 & \text{si } 3 \le x \end{cases}$ verifica:

(a) No es función de distribución de una variable aleatoria discreta ni continua.*
(b) Es función de distribución de una variable aleatoria discreta.
(c) Es función de distribución de una variable aleatoria continua.

Solución
F empieza en 0, termina en 1 y es creciente, pero no es escalonada ni continua.



Sea $X \sim U(a, b)$, con $a < b$.

(a) Si $E(X) < 0$, entonces $b < -a$.*
(b) Si $E(X) = 0$, entonces $b \neq -a$.
(c) Si $E(X) > 0$, entonces $b = |a|$.

Solución
La media de una distribución $U(a,b)$ vale $\mu(X) = (a+b)/2$. Por tanto:

- $\mu(X) < 0 \Rightarrow \dfrac{a+b}{2} < 0 \Rightarrow a+b < 0 \Rightarrow b < -a.$
- $\mu(X) = 0 \Rightarrow \dfrac{a+b}{2} = 0 \Rightarrow a+b = 0 \Rightarrow b = -a.$
- $\mu(X) > 0 \Rightarrow \dfrac{a+b}{2} > 0 \Rightarrow a+b > 0 \Rightarrow b > -a.$

———————— • ————————

Una centralita recibe una media de 2 llamadas cada hora. Por tanto, la probabilidad de que en 3 horas reciba entre 3 y 5 llamadas (ambos valores incluídos) vale:

(a) 0.1339
(b) 0.2945
(c) 0.3837*

Solución
Se considera la variable aleatoria:

$$A : \text{“número de llamadas recibidas en 3 horas”} \sim P(3 \cdot 2) = P(6)\,.$$

Se trata de calcular:

$$p(3 \leq X \leq 5) = p(X \leq 5) - p(X < 3) = p(X \leq 5) - p(X \leq 2) \simeq 0.4457 - 0.0620 = 0.3837.$$

———————— • ————————

Sean $X_1, \ldots, X_{100} \sim G(1/2)$ independientes. La media muestral tiene distribución aproximada:

(a) $\overline{X} \approx N(0,1)$
(b) $\overline{X} \approx N\left(1, \sqrt{2}/10\right)$*
(c) $\overline{X} \approx N\left(1/2, 10/\sqrt{2}\right)$

Solución
La media y la varianza de cada variable $X_i \sim G(1/2)$ valen, respectivamente:

$$\mu(X_i) = \frac{1-1/2}{1/2} = 1 \qquad \text{y} \qquad \sigma^2(X_i) = \frac{1-1/2}{(1/2)^2} = 2.$$

Usando el Teorema del Límite Central ($n = 100$ es grande), se verifica:

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \approx N\left(\mu(X_i), \frac{\sigma(X_i)}{\sqrt{n}}\right) = N\left(1, \frac{\sqrt{2}}{\sqrt{100}}\right) = N\left(1, \frac{\sqrt{2}}{10}\right).$$

———————— • ————————

Sea $X \sim B(5,p)$, con p desconocido. Sea $[X_1, X_2]$ una muestra de X. Se toma $\widehat{p} = 3\,X_1 - 2\,X_2$ como estimador de p. El error cuadrático medio de $\widehat{p}$ vale:

(a) $16\,p^2 + 65\,p\,(1-p)$*
(b) $4\,p^2 + 25\,p\,(1-p)$
(c) $5\,p\,(1-p)$

Sea $X \sim N(\mu, \sigma)$, con σ desconocida. Se toma una muestra de X con $\overline{x} = 2.3$ y $s^2 = 1.5$. Se construye un intervalo de confianza para σ^2 y se obtiene:

$$\left(\frac{(n-1)\,s^2}{b}, \frac{(n-1)\,s^2}{a} \right) = (1.2, 3.4)\,.$$

Entonces, la cota de error más ajustada que se puede dar al estimar σ^2 por s^2 vale:

(a) 0.3
(b) 1.1
(c) 1.9*

Se analiza el número de veces que cierto componente está funcionando en 4 observaciones del mismo. Se observa una muestra de tamaño 80 y se quiere estudiar si los datos proceden de una distribución $B(4, p)$, al 5 % de significación. Se realiza el contraste adecuado y resulta:

Nº veces	Frecuencia	p_i	$n \cdot p_i$	$n_i^2 / (n \cdot p_i)$	$(n_i - n \cdot p_i)^2 / (n \cdot p_i)$
0	13	0.2401	19.208	8.7984	2.0064
1	41	0.4116	32.928	51.0508	1.9787
2	23	0.2646	21.168	24.9906	0.1585
≥ 3	3	0.0837	6.696	1.3441	2.0400
	$n = 80$			86.1838	

El p-valor del contraste verifica:

(a) p-valor > 0.025 y se acepta que los datos vengan de una distribución $B(4, p)$.
(b) p-valor < 0.05 y se rechaza que los datos vengan de una distribución $B(4, p)$.*
(c) p-valor > 0.05 y se acepta que los datos vengan de una distribución $B(4, p)$.

Estadística. Examen final. 20-01-2014

Instrucciones:

- Tiempo para esta parte del examen: 3 horas.
- Entregar la teoría y cada problema en hojas separadas.
- Sólo se puede salir al servicio en casos excepcionales y previa autorización de un profesor.
- Las soluciones, las notas y la fecha de revisión del examen se publicarán en el espacio Moodle de la asignatura.

Teoría (10 % de la nota del examen)

Sean A y B dos sucesos independientes. Demostrar que $\overline{A}$ y $\overline{B}$ también son independientes.

Problema 1 (17 % de la nota del examen)

En una red de ordenadores, el tiempo que tarda un mensaje en llegar a su destino es una variable aleatoria, X. La función de densidad de X (en ms), depende de dos parámetros, $\theta \in \mathbb{R}$ y $\beta > 0$:

$$f(x) = \begin{cases} 0 & \text{si } x < \theta \\ k\, e^{-\beta(x-\theta)} & \text{si } \theta \leq x \end{cases}$$

(a) (2 puntos) Demostrar que $k = \beta$.

(b) (2 puntos) Obtener la función de distribución de X.

(c) (2 puntos) Hallar la función de distribución de la variable aleatoria $Y = X - \theta$. ¿Qué modelo de distribución sigue Y?

(d) (2 puntos) Obtener el tiempo medio que tarda un mensaje en llegar a su destino y la varianza de dicho tiempo.

(e) (2 puntos) Si $\theta = \beta = 1$, calcular la probabilidad de que un mensaje tarde entre 5 ms y 8 ms, sabiendo que lleva en la red 7 ms.

Solución

Apartado (a)

Se considera la variable aleatoria:

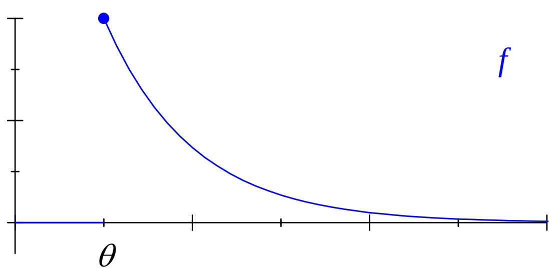
X : "tiempo que tarda un mensaje en llegar a su destino (ms)".

Cualquiera que sea k, la función f es positiva en $\mathbb{R}$ y continua en $\mathbb{R} - \{\theta\}$. Para ser función de densidad, el área bajo la curva debe valer 1:

$$\int_{-\infty}^{\infty} f = \int_{\theta}^{\infty} k\, e^{-\beta(x-\theta)}\, dx = k \left[\frac{e^{-\beta(x-\theta)}}{-\beta} \right]_{\theta}^{\infty} = k \left(0 - \frac{1}{-\beta} \right) = \frac{k}{\beta} = 1 \Rightarrow k = \beta.$$

La expresión de f y su gráfica son:

$$f(x) = \begin{cases} 0 & \text{si } x < \theta \\ \beta\, e^{-\beta(x-\theta)} & \text{si } \theta \leq x \end{cases}$$



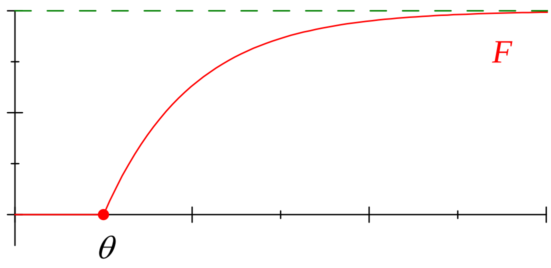
Apartado (b)

Si $x < \theta$, entonces $F(x) = 0$. Si $\theta \leq x$, entonces $F(x)$ vale:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{\theta}^{x} \beta\, e^{-\beta(t-\theta)}\, dt = \left[-e^{-\beta(t-\theta)}\right]_{\theta}^{x} = -e^{-\beta(x-\theta)} - (-1)\,.$$

La expresión de F y su gráfica son:

$$F(x) = \begin{cases} 0 & \text{si } x < \theta \\ 1 - e^{-\beta(x-\theta)} & \text{si } \theta \leq x \end{cases}$$



Apartado (c)

Se trata de obtener la función de distribución de $Y = X - \theta$:

$$F_Y(y) = p(Y \leq y) = p(X - \theta \leq y) = p(X \leq y + \theta) = F_X(y + \theta) =$$
$$= \begin{cases} 0 & \text{si } y + \theta < \theta \\ 1 - e^{-\beta(y+\theta-\theta)} & \text{si } \theta \leq y + \theta \end{cases} = \begin{cases} 0 & \text{si } y < 0 \\ 1 - e^{-\beta\, y} & \text{si } 0 \leq y \end{cases}$$

Esta función de distribución corresponde a $Y \sim \mathrm{Exp}(\beta)$, como se comprueba calculando su derivada.

Apartado (d)

Se trata de calcular la media y la varianza de X. Teniendo en cuenta que $X = Y + \theta$ y que $Y \sim \mathrm{Exp}(\beta)$, se verifica:

$$\mu(X) = \mu(Y + \theta) = \mu(Y) + \theta = \frac{1}{\beta} + \theta,$$
$$\sigma^2(X) = \sigma^2(Y + \theta) = \sigma^2(Y) = \frac{1}{\beta^2}.$$

Como alternativa, se puede calcular la media de X directamente:

$$\mu(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{\theta}^{\infty} x\, \beta\, e^{-\beta(x-\theta)}\, dx.$$

Haciendo el cambio de variable $t = \beta(x - \theta)$, resulta:

$$t = \beta(x - \theta) \Rightarrow x = \frac{t}{\beta} + \theta \Rightarrow dx = \frac{1}{\beta}\, dt,$$
$$\mu(X) = \int_0^{\infty} \left(\frac{t}{\beta} + \theta\right) \beta\, e^{-t}\, \frac{1}{\beta}\, dt = \frac{1}{\beta} \int_0^{\infty} t\, e^{-t}\, dt + \theta \int_0^{\infty} e^{-t}\, dt = \frac{1}{\beta}\, \Gamma(2) + \theta\, \Gamma(1) =$$
$$= \frac{1}{\beta}\, 1! + \theta\, 0! = \frac{1}{\beta}\, 1! + \theta\, 0! = \frac{1}{\beta} + \theta.$$

Análogamente, la media de X^2 vale:

$$\mu\left(X^2\right) = \int_{\theta}^{\infty} x^2\, \beta\, e^{-\beta(x-\theta)}\, dx = \int_0^{\infty} \left(\frac{t}{\beta} + \theta\right)^2 \beta\, e^{-t}\, \frac{1}{\beta}\, dt =$$
$$= \frac{1}{\beta^2} \int_0^{\infty} t^2\, e^{-t}\, dt + \frac{2\,\theta}{\beta} \int_0^{\infty} t\, e^{-t}\, dt + \theta^2 \int_0^{\infty} e^{-t}\, dt = \frac{1}{\beta^2}\, \Gamma(3) + \frac{2\,\theta}{\beta}\, \Gamma(2) + \theta^2\, \Gamma(1) =$$
$$= \frac{1}{\beta^2}\, 2! + \frac{2\,\theta}{\beta}\, 1! + \theta^2\, 0! = \frac{2}{\beta^2} + \frac{2\,\theta}{\beta} + \theta^2.$$

Por tanto, la varianza de X vale:

$$\sigma^2(X) = \mu(X^2) - \mu(X)^2 = \frac{2}{\beta^2} + \frac{2\,\theta}{\beta} + \theta^2 - \left(\frac{1}{\beta} + \theta\right)^2 =$$
$$= \frac{2}{\beta^2} + \frac{2\,\theta}{\beta} + \theta^2 - \frac{1}{\beta^2} - \frac{2\,\theta}{\beta} - \theta^2 = \frac{1}{\beta^2}.$$

Apartado (e)

En este caso, la función de distribución es:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - e^{-(x-1)} & \text{si } 1 \le x \end{cases}$$

Se trata de calcular:

$$p(5 \le X \le 8 \mid X \ge 7) = \frac{p(\{5 \le X \le 8\} \cap \{X \ge 7\})}{p(X \ge 7)} = \frac{p(7 \le X \le 8)}{1 - p(X < 7)} = \frac{F(8) - F(7)}{1 - F(7)} =$$
$$= \frac{\left(1 - e^{-7}\right) - \left(1 - e^{-6}\right)}{1 - (1 - e^{-6})} = \frac{e^{-6} - e^{-7}}{e^{-6}} = 1 - \frac{1}{e} \simeq 1 - 0.36788 = 0.63212.$$

Problema 2 (17 % de la nota del examen)

El tiempo que un servidor tarda en procesar un trabajo que le llega desde una terminal, en minutos, es una variable aleatoria, X, con distribución exponencial de media 30 segundos.

(a) (1.5 puntos) Calcular la probabilidad de que el tiempo que tarda este servidor en procesar un trabajo sea al menos 1 minuto.

Supongamos que se envían trabajos desde distintas terminales en momentos del día donde se puede considerar que la carga del servidor es similar.

(b) (2 puntos) Obtener razonadamente la distribución de probabilidad de la variable que mide el tiempo total que tardan en ser procesados 2 trabajos. Calcular la probabilidad de que el tiempo que tarden en procesarse estos trabajos sea de al menos 2 minutos.

(c) (2 puntos) Si se envían varios trabajos al servidor, calcular la probabilidad de se procesen 3 o más trabajos antes de que un trabajo tarde en procesarse al menos 1 minuto.

(d) (2 puntos) Si se envían 5 trabajos, obtener la probabilidad de que al menos 2 trabajos tarden menos de 1 minuto cada uno en ser procesados por el servidor.

(e) (2.5 puntos) Si se envían 100 trabajos al servidor, calcular la probabilidad de 80 o menos tarden en procesarse cada uno menos de 1 minuto.

Solución

Apartado (a)

Se considera la variable aleatoria:

$$X : \text{“tiempo que tarda el servidor en procesar un trabajo (min)”} \sim \mathrm{Exp}(\beta)$$

Se sabe que $\mu(X) = 1/\beta = 1/2$, luego $\beta = 2$.

Sea X≈exp(β), E[X]=1/β= 30/60 minutos =1/2 => β=2

a) $P(X\geq 1) = 1 - F(1) = e^{-2} = 0´1353$,, $F(1)=\int_0^1 2e^{-2x}\,dx=1-e^{-2}=0´1353$
b) El tiempo que tarda en procesar dos trabajos será el tiempo que tarda en procesar el primer trabajo ($X_1\approx exp(2)$) más el tiempo que tarda en procesar el segundo ($X_2\approx exp(2)$), y estos los podemos considerar independientes. $T\approx X_1+X_2\approx\gamma(2,2)$

$$P(T\geq 2)=\int_2^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}dx = 4\int_2^\infty x\, e^{-2x}dx =$$

= (resolviendo por partes, u=x y dv=$e^{-2x}dx$) = $5e^{-5}$= 0´0916

c) Sea Y≈números de trabajos que se procesan antes de que un trabajo tarde en procesarse al menos un minuto≈G(p) con p=0´1353

$$P(Y\geq 3)=(1-p)^3=0´6465$$

d) Sea Z≈números de trabajos que tardan menos un minuto en procesarse≈B(5,1-p) con p=0´1353

$$P(Z\geq 2)=1-P(Z\leq 1)=1-0´135^5-5*0´1353^4*0´8647=0´9985$$

e) Sea N≈números de trabajos de los 100 que tardan menos un minuto en procesarse≈B(100,0´8647), al ser n grande aproximamos por una $N(np,\sqrt{np(1-p)})\approx N(86´47,3´42)$

$$P(N(86´47,3´42)\leq 80))= P(N(0,1)\leq\frac{80-86´47}{3.42})=$$

$$=P(N(0,1)\leq -1´89)= 1-P(N(0,1)\leq 1´89)=0´0294$$

Problema 3 ($16\,\%$ de la nota del examen)

El tiempo que transcurre desde que cierto modelo de servidor se estropea hasta que es reparado, en horas, es una variable aleatoria con densidad:

$$f(x) = \begin{cases} 0 & \text{si } x < 3 \\ \alpha\, 3^{\alpha}/x^{\alpha+1} & \text{si } x \geq 3 \end{cases} \qquad \alpha > 2.$$

Se toma una muestra aleatoria simple de dicho tiempo, con tamaño $n = 256$, media $4.6\,\mathrm{h}$, cuasivarianza $4\,\mathrm{h}^2$ y $\sum_{i=1}^{n} \ln(x_i) = 373.396$. Se pide:

(a) (2 puntos) Obtener el estimador y la estimación de α por el método de los momentos.

(b) (2.5 puntos) Obtener el estimador y la estimación de α por máxima verosimilitud.

(c) (2.5 puntos) Construir razonadamente, a partir del pivote correspondiente, un intervalo, al $90\,\%$ de confianza, para el tiempo medio que transcurre hasta que el servidor es reparado.

(d) (3 puntos) En este tipo de servidor, el tiempo de ejecución de un proceso, en segundos, sigue una distribución normal con desviación típica $0.2\,\mathrm{s}$. Se han hecho ciertos cambios en el servidor y se sospecha que la variabilidad en el tiempo de ejecución ha aumentado. Se toma una muestra de tamaño 20, resultando que la cuasivarianza muestral es $0.07\,\mathrm{s}^2$. A la vista de los datos y con un nivel de significación del $5\,\%$ ¿es cierta la sospecha que se tiene? Obtener el p-valor.

Apellidos:	
Nombre:	Titulación:

Estadística. Examen final. **20-01-2014**

Parte con ordenador.

- **Tiempo: 50 minutos**
- **Valor: 20%**

MODELOS A, B

Dos aseguradoras A y B, de fuerte implantación, han realizado un estudio sobre los costes de reparación de los vehículos asegurados en cierta ciudad. Los datos se han recogido durante los días laborables de un año (252) y están contenidos en las variables X(A), X(B), F(A) y F(B) del fichero Enero2014.sf3.

La variable X(A) contiene el número de vehículos asegurados por A que llegan diariamente a los talleres para ser reparados. Para esta variable:

(a) (1.5 puntos) Obtener los valores de los cuartiles primero y tercero, hallar los límites de admisión e indicar si la muestra contiene datos atípicos. En caso afirmativo, dar el valor de uno de ellos.

Observando la tabla de estadísticos, se obtiene que:

Summary Statistics for X(A)

Count	252
Average	7,84524
Variance	7,99587
Coeff. of variation	36,0435%
Minimum	2,0
Maximum	16,0
Range	14,0
Lower quartile	6,0
Upper quartile	10,0
Interquartile range	4,0
Skewness	0,540269

$Q_1 = 6;\ Q_3 = 10$. Por tanto los límites de admisión son:

$$L_1 = Q_1 - \frac{3}{2}R_Q = 6 - \frac{3}{2}4 = 0;$$

$$L_3 = Q_3 + \frac{3}{2}R_Q = 10 + \frac{3}{2}4 = 16$$

Como los valores mínimo y máximo son 2 y 16, no hay datos atípicos en la muestra.

Modelo B: $Q_1 = 5;\ Q_3 = 9$

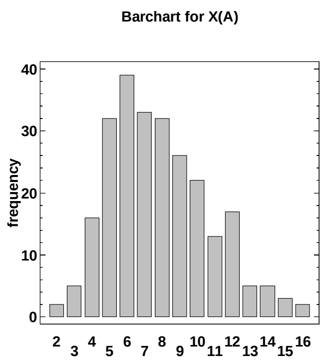
$$L_1 = Q_1 - \frac{3}{2}R_Q = 5 - \frac{3}{2}4 = -1;$$

$$L_3 = Q_3 + \frac{3}{2}R_Q = 9 + \frac{3}{2}4 = 15$$

Como el mínimo es 1 y el máximo 14, tampoco hay datos atípicos.

(b) (1.5 puntos) Ajustar razonadamente (es necesario escribir los motivos por los que se hace la hipótesis) un modelo de probabilidad a la variable X(A) y estimar sus parámetros a mano (utilizando el método de los momentos) y con Statgraphics.

La variable X(A) es discreta (la muestra toma valores enteros entre 2 y 16). Si dibujamos el diagrama de barras



podemos observar que es asimétrica a la derecha y tiene un perfil que podría corresponder a una binomial con $p < 0'5$ o a una Poisson. Ahora bien, sabemos que X(A) cuenta el número de veces que ocurre un suceso en un período de tiempo y si observamos los estadísticos que hemos obtenido en el apartado anterior vemos que su media y su varianza son similares, por lo que proponemos un modelo de Poisson. Como la media muestral es 7'84524, utilizando el método de los momentos y que

$E\left[X(A)\right] = \lambda$,

resulta que una estimación de λ es 7'84524 y el modelo propuesto es X(A) ~ $P(7'84524)$. Si contrastamos con Statgraphics obtenemos:

Chi-Square = 12,5712 with 12 d.f. P-Value = 0,400964, por lo que aceptamos la hipótesis.

Para el Modelo B:

Para el Modelo B: X(B) ~ $P(6'9444)$ y P-Value = 0,320507 también se acepta.

(c) (1 punto) Utilizando los parámetros obtenidos por Statgraphics, calcular la probabilidad de que un día cualquiera lleguen a los talleres más de 10 vehículos.

Suponiendo que X(A) ~ $P(7'84524)$, se tiene que: $P(X(A) > 10) = 0{,}169055$.

Modelo B: X(B) ~ $P(6'9444)$, se tiene que: $P(X(B) > 6) = 0{,}541972$.

(d) (1.5 puntos) Obtener un intervalo de confianza al 92% para el número medio de vehículos que llegan a los talleres para ser reparados. Es necesario detallar todas las hipótesis necesarias para construir el intervalo. ¿Se puede admitir que llegan a los talleres una media de 8 vehículos diarios para reparar?

Según hemos comprobado en el apartado (b), la variable X(A) no es normal, pero la muestra es suficientemente grande (252) y podemos aplicar el teorema central del límite y obtener un intervalo aproximado para la media. Calculándolo con Statgraphics, obtenemos:

92,0% confidence interval for mean: 7,84524 +/- 0,313115 [7,53212; 8,15835]. Como 8 pertenece al intervalo, con ese nivel de confianza se puede admitir que llegan 8 vehículos a los talleres.

Para el modelo B, usando un nivel de confianza del 90%:

La variable tampoco se puede considerar normal. Razonando análogamente se obtiene:

90,0% confidence interval for mean: 6,94444 +/- 0,263005 [6,68144; 7,20745]

Como 8 no pertenece al intervalo no se puede admitir que llega una media de 8 vehículos diarios.

La aseguradora A sospecha sobre la facturación de los talleres, por lo que hace un estudio sobre el importe diario que paga a dichos talleres. La variable F(A) recoge los valores de estos importes, en euros. No necesariamente todos los talleres de la ciudad son muestreados todos los días. La aseguradora A afirma que el importe medio diario que paga a los talleres de esa ciudad es de 4000€ con una desviación típica menor de 100€, lo cual le hace sospechar la existencia de algún acuerdo entre los talleres. Utilizando un nivel de significación del 4% (En el Modelo B 6%):

(e) (2.5 puntos) Contrastar las afirmaciones de la aseguradora. Es necesario justificar todas las hipótesis necesarias para efectuar el estudio, así como el planteamiento de los contrastes necesarios para realizarlo.

Debemos plantear dos contrastes:

$$\begin{cases} \mathrm{H}_0: \mu = 4000 \\ \mathrm{H}_1: \mu \neq 4000 \end{cases}; \quad \begin{cases} \mathrm{H}_0: \sigma = 100 \\ \mathrm{H}_1: \sigma < 100 \end{cases} \qquad \begin{cases} \mathrm{H}_0: \mu = 3800 \\ \mathrm{H}_1: \mu \neq 3800 \end{cases}; \quad \begin{cases} \mathrm{H}_0: \sigma = 80 \\ \mathrm{H}_1: \sigma < 80 \end{cases}$$

Miramos en primer lugar si la variable F(A) es normal, porque en caso de no serlo, no podríamos hallar el intervalo para la varianza.

A la vista de p-valor, asumimos normalidad.

Resolvemos los contrastes con Statgraphics:

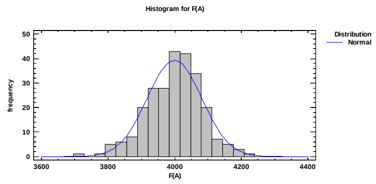
Fitted Distributions

Normal
mean = 3999,34
standard deviation = 81,9485

Goodness-of-Fit Tests for F(A)

Kolmogorov-Smirnov Test

	Normal
DPLUS	**0,0317586**
DMINUS	**0,0587412**
DN	**0,0587412**
P-Value	**0,35232**



t-test
Null hypothesis: mean = 4000,0
Alternative: not equal

Computed t statistic = -0,127658
P-Value = 0,898521
Do not reject the null hypothesis for alpha = 0,04.

chi-square test
Null hypothesis: sigma = 100,0
Alternative: less than

Computed chi-square statistic = 168,561
P-Value = 0,0000168268
Reject the null hypothesis for alpha = 0,04.

Por tanto, con ese nivel de confianza, admitimos las sospechas de la aseguradora.

Para el modelo B:

La variable F(B) es normal (P-Value=0,973473). Resolvemos los contrastes:

t-test
Null hypothesis: mean = 3800,0
Alternative: not equal

Computed t statistic = 1,97745
P-Value = 0,0490852
Reject the null hypothesis for alpha = 0,06.

chi-square test
Null hypothesis: sigma = 80,0
Alternative: less than

Computed chi-square statistic = 299,339
P-Value = 0,980391
Do not reject the null hypothesis for alpha = 0,06.

Con ese nivel de significación solo admitimos que la desviación típica es menor que 80€.

(f) (2 puntos) Contrastar si hay diferencias significativas en los importes pagados por las aseguradoras A y B a los talleres de reparación de esa ciudad. Es necesario justificar las hipótesis necesarias para efectuar el estudio, así como el planteamiento de los contrastes necesarios para realizarlo.

Nos piden contrastar:

$$\begin{cases} H_0: \mu_{F(A)} - \mu_{F(B)} = 0 \\ H_1: \mu_{F(A)} - \mu_{F(B)} \neq 0 \end{cases}$$

En primer lugar, observamos que las muestras no son pareadas porque, tal y como dice en el enunciado, no todos los talleres reciben coches de ambas aseguradoras. Por tanto estamos en el caso de muestras independientes.

Necesitamos, por tanto, saber si podemos considerar que las varianzas son iguales, para lo cual debemos contrastar si la variable F(B) es normal. Utilizamos Statgraphics:

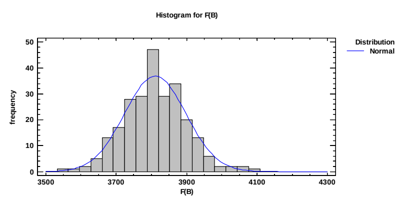
Fitted Distributions

Normal
mean = 3810,88
standard deviation = 87,3645

Goodness-of-Fit Tests for F(B)

Kolmogorov-Smirnov Test

	Normal
DPLUS	0,0304645
DMINUS	0,0218801
DN	0,0304645
P-Value	0,973473



A la vista del gráfico y el p-valor, admitimos que F(B) es normal.

Contrastamos ahora la igualdad de varianzas:

$$\begin{cases} H_0 : \dfrac{\sigma^2_{F(A)}}{\sigma^2_{F(B)}} = 1 \\ H_1 : \dfrac{\sigma^2_{F(A)}}{\sigma^2_{F(B)}} \neq 1 \end{cases}$$

F-test to Compare Standard Deviations
Null hypothesis: sigma1 = sigma2
Alt. hypothesis: sigma1 NE sigma2
F = 0,879858   P-value = 0,311273
Do not reject the null hypothesis for alpha = 0,04.

Por tanto, admitimos que las varianzas son iguales.

Contrastamos la igualdad de medias:

t test to compare means
Null hypothesis: mean1 = mean2
Alt. hypothesis: mean1 NE mean2
assuming equal variances: t = 24,9757   P-value = 0
Reject the null hypothesis for alpha = 0,04.

Por tanto no admitimos, con ese nivel de significación que ambas aseguradoras pagan lo mismo.

Este apartado es igual para el modelo B, con α=0.06.