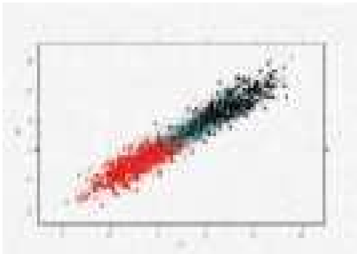
ESTADÍSTICA I (SISTEMAS)

Profesores: Hilario Navarro. Jorge Martín



DEPARTAMENTO DE ESTADÍSTICA,
INVESTIGACIÓN OPERATIVA Y CÁLCULO
NUMÉRICO

Soluciones a los problemas del examen.
Convocatoria de septiembre. Curso 2002-2003

Problema 1. Se está realizando un estudio sobre los fallos de un dispositivo electrónico. Este elemento se puede montar en dos posiciones diferentes y hay cuatro tipos de fallos posibles. Un muestreo aleatorio proporciona la siguiente distribución de frecuencias:

	Tipo de fallo			
Posición de Montaje	A	B	C	D
1	14	18	8	20
2	6	12	12	10

¿Concluiría que el tipo de fallo es independiente de la posición de montaje?

Datos auxiliares:
$t_{3;0.025} = 3.182$, $\chi_{8;0.05} = 15.507$, $\chi_{3;0.05} = 7.815$, $z_{0.05} = 1.645$

Solución

Rechazaremos la hipótesis nula (*Posición de Montaje* independiente del *Tipo de Fallo*), con un nivel de sgnificación α, si

$$\sum_{j=1}^{4}\sum_{i=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} > \chi^2_{(2-1)(4-1);\alpha}$$

El enunciado nos proporciona la frecuencia observada (o_{ij}); multiplicando las correspondientes frecuencias marginales y dividiendo por el tamaño de la muestra obtenemos la frecuencia esperada bajo la hipótesis nula (e_{ij}):

	Tipo de fallo				
Posición de Montaje	A	B	C	D	
1	12	18	12	18	60
2	8	12	8	12	40
	20	30	20	30	100

Con estos datos podemos realizar la operación $(o-e)^2/e$ para cada celda de la tabla, resultando:

	Tipo de fallo			
Posición de Montaje	A	B	C	D
1	4/12	0	16/12	4/18
2	4/8	0	16/8	4/12

Entonces,

$$\sum_{j=1}^{4}\sum_{i=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \frac{4+16+4}{12} + \frac{16+4}{8} + \frac{4}{18} = \frac{144+180+16}{72} = \frac{340}{72} \simeq 4.7$$

Como $4.7 < \chi^2_{3;0.05} = 7.815$, concluimos que, con un nivel de significación $\alpha = 0.05$, la condición de **independencia** es **aceptable**.

$\square$

Problema 2. Cierto aparato registra el nivel de saturación de la red eléctrica en una comarca. El error relativo porcentual de la medida dada por el aparato es una variable aleatoria continua X con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1-x)^3 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Determinar:

(a) La función de densidad de la variable X.

(b) La probabilidad de que una medida registrada por el aparato tenga un error entre el 0.1 % y el 0.2 %.

(c) El error relativo medio.

Solución

(a) La representación gráfica de la función de distribución $F(x)$ es la que aparece en la figura 1.

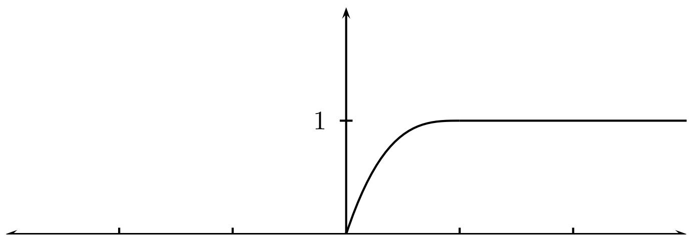


Figura 1: Función de distribución de la variable aleatoria X

Puesto que la variable X es continua, la función de densidad se obtiene derivando la de distribución. Dicha función viene dada por

$$f(x) = F'(x) = \begin{cases} 0 & \text{si } x < 0 \\ 3(1-x)^2 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si } x > 1 \end{cases}$$

Su representación gráfica es la que aparece en la figura 2.

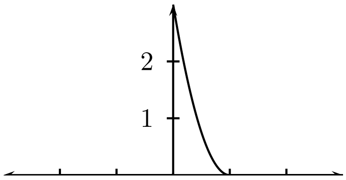


Figura 2: Función de densidad de la variable aleatoria X

(b) El error de medida está entre el 0.1 % y el 0.2 % cuando $0.1 \leq X \leq 0.2$. Por tanto, la probabilidad pedida será

$$P(0.1 \leq X \leq 0.2) = \int_{0.1}^{0.2} f(x)\,dx = 3\int_{0.1}^{0.2} (1-x)^2\,dx = 0.217.$$

Esta probabilidad es el área sombreada de la figura 3.

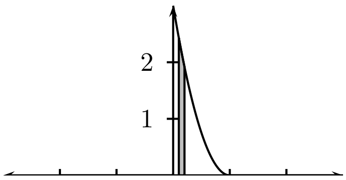


Figura 3: $P(0.1 \leq X \leq 0.2)$

Un modo alternativo de llegar al mismo resultado es haciendo uso de la función de distribución dada en el enunciado del problema.

$$P(0.1 \leq X \leq 0.2) = \int_{0.1}^{0.2} f(x)\,dx = \int_{-\infty}^{0.2} f(x)\,dx - \int_{-\infty}^{0.1} f(x)\,dx$$

$$= F(0.2) - F(0.1) = 1 - (1-0.2)^3 - (1-(1-0.1)^3) = 0.217.$$

(c) Finalmente, en este apartado nos están pidiendo la media de la variable aleatoria X que mide el error, la cual viene dada por

$$E\{X\} = \int_{-\infty}^{\infty} x f(x)\,dx = 3\int_0^1 x(1-x)^2\,dx = 3\left(\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4}\right)\Bigg]_0^1 = \frac{1}{4}$$

$\square$

Problema 3. La cantidad —en Kg— de cereal cosechada por m^2 en una región es una variable aleatoria con distribución *normal*. En 25 localizaciones elegidas al azar se obtuvo que la cantidad media cosechada por m^2 fue de 18.5 kg con una cuasivarianza de 1 kg. Contrastar la hipótesis de que la cantidad media por m^2 es de 18 kg, frente a la alternativa de que es mayor. Tómese un nivel de significación de $\alpha = 0.1$.

Datos auxiliares: $z_{0.01} = 2.33$, $t_{24;0.01} = 2.192$, $t_{25;0.01} = 2.485$

Solución

El contraste de hipótesis sobre la cantidad media μ de cereal cosechado por m^2 viene dado por

$$H_0 : \mu = 18 \qquad\qquad H_1 : \mu > 18.$$

Dado que la cantidad recolectada por m^2 es una variable aleatoria con distribución normal de varianza desconocida, el estadístico de contraste que debemos emplear es

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

donde

- μ_0: la media poblacional bajo H_0. En este caso $\mu_0 = 18$.
- $\bar{X}$: media muestral. Para la muestra extraída $\bar{X} = 25$.
- S^2: cuasivarianza muestral. Para la muestra extraída $S^2 = 1$.
- n: tamaño muestral. En este caso $n = 25$.

La región crítica del test viene dada por

$$\mathcal{R} = \left\{ \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{n-1;\alpha} \right\},$$

siendo α el nivel de significación y $t_{n-1;\alpha}$ el cuantil $1 - \alpha$ de una distribución t con $n - 1$ grados de libertad, es decir, el valor de la distribución que verifica que $P(t_{n-1} \leq t_{n-1;\alpha}) = 1 - \alpha$.

En este caso, para un tamaño muestral $n = 25$ y un nivel de significación $\alpha = 0.01$, la región crítica —zona sombreada de la figura 4— viene dada por

$$\mathcal{R} = \left\{ \frac{\bar{X} - 18}{S/5} > t_{24;0.01} = 2.192 \right\}.$$

Para la muestra extraída se obtendrá que $\dfrac{\bar{X} - 18}{S/5} = \dfrac{18.5 - 18}{1/5} = 2.5 \in \mathcal{R}$, lo cual conduce a rechazar la hipótesis nula de que la cantidad media de cereal cosechada por m^2 es de 18 Kg.

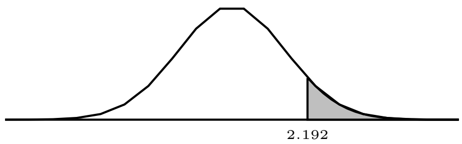


Figura 4: Función de densidad de una t_{24}

□

Problema 4. Para cada una de las condiciones que se indican a continuación, represente una nube de puntos (X, Y) que sea compatible con ella:

- Covarianza negativa.
- Pendiente de la recta de regresión de Y sobre X positiva.
- Correlación próxima a 1.
- Correlación nula.

(Nota: Justifique brevemente cada representación)

Solución

- *Breve justificación*: La relación entre **covarianza**, **coeficiente de correlación lineal** y **pendiente de la recta de regresión** de Y sobre X se pone de manifiesto en las siguientes igualdades:

$$Pendiente = \frac{cov_{x,y}}{v_x} = r\sqrt{\frac{v_y}{v_x}}$$

Por tanto, dichos elementos tendrán siempre el mismo signo (gráficas de la figura 5).

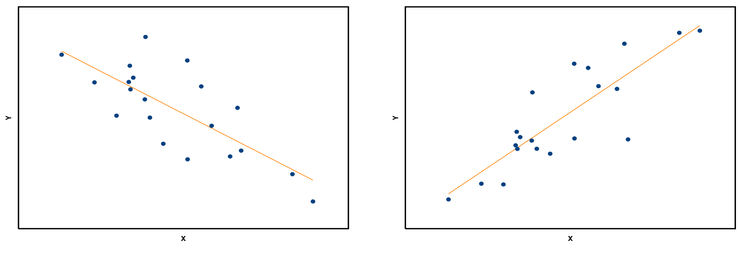


Figura 5: Covarianza negativa. Pendiente positiva

Además, el valor del **coeficiente de correlación** se refleja en la **varianza residual** según indica la siguiente expresión:

$$varianza\ residual = v_y \left(1 - r^2\right) \ .$$

Así, el caso de correlación próxima a 1 se corresponde con un valor pequeño para el cociente $\dfrac{varianza\ residual}{v_y}$ (gráfica izquierda de la figura 6), mientras que un coeficiente de correlación nulo supone una varianza residual cercana a su valor máximo, que es v_y (gráfica derecha de la figura 6).

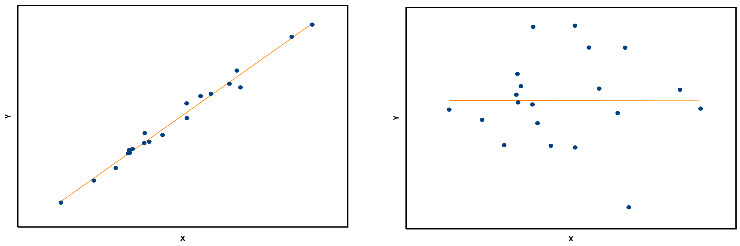


Figura 6: Correlación próxima a 1. Correlación nula

$\square$