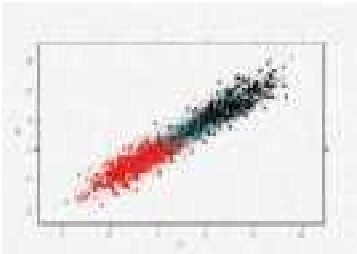
ESTADÍSTICA (SISTEMAS)

Profesores: Hilario Navarro. Jorge Martín



DEPARTAMENTO DE ESTADÍSTICA, INVESTIGACIÓN OPERATIVA Y CÁLCULO NUMÉRICO

Soluciones a los problemas del examen.
Convocatoria de septiembre. Curso 2003-2004

Problema 1. La tabla siguiente muestra el número de artículos **defectuosos** y **aceptables** que se obtienen **antes** y **después** de introducir una modificación en un proceso de fabricación

	Defectuoso	*Aceptable*
Antes	14	46
Después	6	34

Utilice un contraste χ^2 para decidir si el cambio es significativo ($\alpha = 0.01$).

Datos auxiliares: $\chi^2_{2;0.01} = 9.210$, $\chi^2_{4;0.01} = 13.277$, $\chi^2_{1;0.01} = 6.635$

Solución

Para decidir si el cambio observado en la distribución de "Aceptable/Defectuoso" es significativo —aplicando un contraste del tipo χ^2—, tendremos que comparar el valor del estadístico

$$\sum_{j=1}^{2}\sum_{i=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

con el de $\chi^2_{(2-1)(2-1);\alpha}$. 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}):

	Defectuoso	*Aceptable*	
Antes	12	48	60
Después	8	32	40
	20	80	100

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

	Defectuoso	*Aceptable*
Antes	$4/12$	$4/48$
Después	$4/8$	$4/32$

Entonces,

$$\sum_{j=1}^{2}\sum_{i=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \frac{32+8+48+12}{96} = \frac{100}{96} \simeq 1.04$$

Como $1.04 < \chi^2_{1;0.05} = 6.635$, concluimos que, con un nivel $\alpha = 0.05$, el cambio no es estadísticamente significativo y, por tanto, es razonable aceptar que la modificación introducida en el proceso de fabricación no tiene el efecto deseado. $\square$

Problema 2. En la fabricación de cierta clase de dispositivos electrónicos se puede emplear uno cualquiera de dos procesos: A y B. El tiempo de vida —en miles de horas— de los dispositivos obtenidos del proceso A es una variable aleatoria X con densidad

$$f(x) = \begin{cases} e^{1-x} \;, & x \geq 1 \\ 0 \;, & \text{en otro caso.} \end{cases}$$

Por otro lado, el tiempo de vida de los dispositivos fabricados con el B es una variable aleatoria Y con densidad

$$g(y) = \begin{cases} \frac{3}{y^4} \;, & y \geq 1 \\ 0 \;, & \text{en otro caso.} \end{cases}$$

Si ambos procesos se comparan por la vida media de los dispositivos que producen, ¿cuál utilizaría?

Solución

Puesto que el criterio de comparación es la vida media de los dispositivos, calcularemos la esperanza de ambas variables.

Para la variable aleatoria X se tiene que

$$E\{X\} = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{1}^{\infty} x e^{1-x}\, dx = -x e^{1-x}\Big]_1^{\infty} + \int_1^{\infty} e^{1-x} dx = 2.$$

Para la variable aleatoria Y la media es

$$E\{Y\} = \int_{-\infty}^{\infty} y f(y)\, dy = 3\int_1^{\infty} y \frac{dy}{y^4} = 3\int_1^{\infty} y^{-3}\, dy = \frac{3y^{-2}}{-2}\Big]_1^{\infty} = \frac{3}{2}.$$

Puesto que $E\{X\} > E\{Y\}$, utilizaremos el proceso A para la fabricación de los dispositivos.

$\square$

Problema 3. A fin de comparar las medias de dos poblaciones con distribución normal de varianzas 2 y 3, respectivamente, se obtuvieron dos muestras aleatorias de tamaños 10 y 9 dadas por:

Pobl. 1	5	8	7	8.5	6.5	8	6	7	6	8
Pobl. 2	7	6	8	5.5	6.5	5	5	6	5	

Contrastar la hipótesis de igualdad de las medias de las dos poblaciones. Tómese un nivel de significación de $\alpha = 0.05$.

Datos auxiliares: Si F es la función de distribución de una $N(0,1)$: $F(-1.645) = 0.05$, $F(-2.57) + 1 - F(2.57) = 0.01$, $F(1.96) = 0.975$

Solución

Dado que se pide contrastar la igualdad de medias de dos poblaciones, la hipótesis nula y la alternativa vienen dadas por

$$H_0 : \mu_1 = \mu_2 \qquad\qquad H_1 : \mu_1 \neq \mu_2$$

siendo μ_1 y μ_2 las medias poblacionales.

Puesto que las poblaciones son normales de varianzas conocidas, el estadístico de contraste que emplearemos es

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

con σ_1^2 y σ_2^2 las varianzas de ambas poblaciones.

La distribución del estadístico anterior es $N(0, 1)$. Los ingredientes muestrales que aparecen en la expresión que lo define son:

- n_1: tamaño de la muestra procedente de la población 1.
- n_2: tamaño de la muestra procedente de la población 2.
- $\bar{X}_1$: media de la muestra extraída de la primera población.
- $\bar{X}_1$: media de la muestra extraída de la segunda población.

Puesto que el contraste es bilateral, la región crítica del test viene dada por

$$\mathcal{R} = \left\{ \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_{\alpha/2} \right\},$$

donde el punto crítico $z_{\alpha/2}$ es el valor de una variable aleatoria Z con distribución $N(0, 1)$ que verifica que $P(Z > z_{\alpha/2}) = \alpha/2$.

En nuestro caso, con tamaños muestrales de $n_1 = 10$ y $n_2 = 9$, para un nivel de significación de $\alpha = 0.05$, el punto crítico será $z_{\alpha/2} = 1.96$, dado que se cumple que $P(Z > 1.96) = 1 - F(1.96) = 0.025$. Por tanto, la región crítica —zona sombreada de la figura 1— vendrá determinada por

$$\mathcal{R} = \left\{ \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > 1.96 \right\}.$$

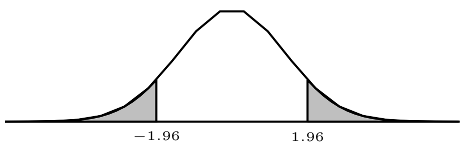


Figura 1: Región crítica para $\alpha = 0.05$

De los datos proporcionados, se obtendrá que

- $\bar{X}_1 = \dfrac{5+8+\cdots+6+8}{10} = 7.$
- $\bar{X}_2 = \dfrac{7+6+\cdots+6+5}{9} = 6.$

Consecuentemente, para la muestra extraída se tiene que

$$\frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{7-6}{\sqrt{\frac{2}{10} + \frac{3}{9}}} = \frac{\sqrt{30}}{4}$$

valor que no pertenece a la región crítica $\mathcal{R}$; con lo cual se acepta la hipótesis de igualdad de las medias de ambas poblaciones.

□

Problema 4. Con el fin de estudiar la relación existente entre las columnas X e Y de una base de datos que contiene 10 registros del par (X, Y), se han realizado los siguientes cálculos: $\sum x_i = 55$, $\sum y_i = 61$, $\sum x_i^2 = 385$, $\sum y_i^2 = 545$, $r = -0.875$.

1. Resuma la correspondiente nube de puntos mediante la recta de regresión de Y sobre X.
2. ¿Cómo valoraría la bondad de dicho resumen? (Justifique —sin demostraciones— su elección)

Datos auxiliares: De los datos del enunciado se obtiene que la relación entre desviaciones típicas es

$$\sqrt{v_y} = 1.448\sqrt{v_x}$$

Solución

(a) La relación entre el **coeficiente de correlación lineal** y la **pendiente de la recta de regresión** de Y sobre X se pone de manifiesto en la siguiente igualdad:

$$Pendiente = r\sqrt{\frac{v_y}{v_x}} \ .$$

Con los datos del enunciado resulta

$$Pendiente = (-0.875) \times (1.448) \simeq -1.27$$

y, dado que

$$\bar{x} = \frac{1}{10}\sum x_i = 5.5 \quad , \quad \bar{y} = \frac{1}{10}\sum y_i = 6.1 \ ,$$

la expresión de la recta de regresión de Y sobre X es

$$y - 6.1 = -1.27\,(x - 5.5)$$

(b) La forma más directa es midiendo del error cuadrático medio, que se conoce como **varianza residual**. Sin embargo, resulta más útil —tiene mayor capacidad de interpretación— el cociente

$$\frac{\text{Varianza residual}}{v_y}$$

que toma valores en el intervalo $[0, 1]$: un valor próximo a 1 refleja un mal ajuste mientras un valor cercano a 0 indica la cualidad contraria. El cálculo de dicho cociente se puede realizar fácilmente, ya que

$$\frac{\text{Varianza residual}}{v_y} = 1 - r^2 \ .$$

En definitiva, podríamos valorar el ajuste simplemente a través del coeficiente r^2. En este caso se obtiene $r^2 \simeq 0.77$, que, en general, se considera como indicador de un ajuste no demasiado bueno.

□