Estadística I
Solución Examen Final- 19 de junio de 2009

Nombre y Apellido:..
Grupo:.................

(1) La siguiente tabla muestra las distribuciones de frecuencias absolutas de la variable `altura` (en metros) de $n = 500$ estudiantes de un centro educativo.

(a) (4.5 puntos) Dibuja el histograma de la variable, utilizando la información de la tabla anterior y describe la forma de la distribución (utiliza la rejilla de la cara posterior de esta página).
Solución GRAPH Bell-shaped.

(b) (1.2 puntos) Calcula las frecuencias relativas.

Intervalo	Frecuencias absolutas	Frecuencias relativas
$[1.59, 1.61)$	11	0.022
$[1.61, 1.63)$	71	0.142
$[1.63, 1.65)$	159	0.318
$[1.65, 1.67)$	177	0.354
$[1.67, 1.69)$	67	0.134
$[1.69, 1.71)$	15	0.030

(c) (2 puntos) ¿Cuántos estudiantes miden más de 1.67m, inclusive?
Solución 82
¿Qué porcentaje de estudiantes miden menos de 1.63m?
Solución 16.4%

d) (2.3 puntos) Se analizó con `R` la variable `altura`, obteniéndose los siguientes resultados,

mean	sd	0%	25%	50%	75%	100%	n
1.6505	0.0202	1.5990	1.6368	1.6508	1.6633	1.7018	500

Usa los resultados anteriores para identificar, o calcular, las siguientes medidas: media, mediana, desviación estándar, varianza, valor mínimo, valor máximo, rango, primer cuartil, tercer cuartil, rango intercuartílico y coeficiente de variación (completa en la cara posterior de esta página).
Solución

$$\begin{array}{ccc}
\bar{x} & = & 1.6505 \text{ (given)} \\
\text{M} & = & 1.6508 \text{ (given)} \\
\text{s} & = & 0.0202 \text{ (given)} \\
s^2 & = & 0.0202^2{=}0.0004 \\
\text{min} & = & 1.5990 \text{ (given)} \\
\text{max} & = & 1.7018 \text{ (given)} \\
\text{range} & = & \text{max} - \text{min}{=}0.1028 \\
Q_1 & = & 1.6368 \text{ (given)} \\
Q_3 & = & 1.6633 \text{ (given)} \\
\text{IQR} & = & Q_3 - Q_1{=}\ 0.0265 \\
\text{CV} & = & \frac{sd}{|\bar{x}|}{=}\ 0.0122
\end{array}$$

(2) Un mayorista de máquinas fotocopiadoras quiere realizar un análisis sobre la posible relación entre el precio de venta, X, y el número de ventas, Y, utilizando la información obtenida de ocho proveedores de dicha marca.

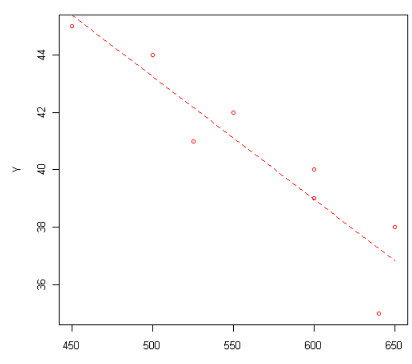
Precio (X)	550	600	640	600	500	650	450	525
Ventas (Y)	42	39	35	40	44	38	45	41

A continuación se muestra la salida de `R` con el modelo de regresión ajustado para estas dos variables.

```
Call: lm(formula = Y ~ X, data = Datos)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2571 -0.6006  0.3836  0.9196  1.1717

Coefficients:
             Estimate     Std.Error    t value Pr(>|t|)
(Intercept)  64.70135     4.02861      16.060  3.7e-06 *** X
-0.04288     0.00709      -6.048 0.000925 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1


Residual standard error: 1.318 on 6 degrees of freedom.
 Multiple R-squared: 0.8591, Adjusted R-squared: 0.8356
F-statistic: 36.58 on 1 and 6 DF,  p-value: 0.000925
```

(a) Representa gráficamente las variables mediante un diagrama de dispersión.

Solución



(b) Obtén, indicando cómo lo haces, el coeficiente de correlación. ¿Crees que existe relación entre el precio y el número de ventas de máquinas fotocopiadoras? ¿De qué tipo y por qué?

Solución

El coeficiente de correlación es la raíz del coeficiente de determinación.

El coeficiente de determinación viene recogido en la salida de R como *Multiple R-squared*. Por tanto:

$$r_{(x,y)} = \sqrt{r^2_{(x,y)}} = \sqrt{0.8591} = -0.9269.$$

Elegimos la raiz negativa, ya que vemos en el diagrama de dispersión que la relación entre las dos variables es negativa. Como es un valor muy grande (en valor absoluto), podemos decir que existe una relación lineal fuerte entre X e Y, y negativa.

(c) Obtén, indicando cómo lo haces, la ecuación de la recta de regresión de Y sobre X y trázala sobre el diagrama de dispersión anterior.

Solución
Los coeficientes de la recta de regresión vienen recogidos en la salida de R en la columna *Estimate* del apartado *Coefficients* como *Intercept* (ordenada en el origen) y x (pendiente).
La ecuación de la recta de regresión es: $y = 64.7014 - 0.0429x$.

(3) Una compañía especializada en instalar calderas estima que las probabilidades de instalar X calderas al mes son las siguientes.

X	0	1	2	3	4	5
probabilidad	0.10	0.13	0.25	0.29	0.16	0.07

(a) Calcular la probabilidad de que tengan que instalar al menos 3 calderas en un mes.

Solución
$P(X \geq 3) = 0.29 + 0.16 + 0.07 = 0.52$

(b) Si en un mes ya se han instalado al menos dos calderas, calcula la probabilidad de que, como mucho, se instalen 4.

Solución

$$P(X \leq 4/X \geq 2) = \frac{P(X \leq 4 \cap X \geq 2)}{P(X \geq 2)} = \frac{0.25 + 0.29 + 0.16}{0.25 + 0.29 + 0.16 + 0.07} = \frac{0.70}{0.77} = 0.9091$$

(c) De 6 meses elegidos al azar, ¿cuál es la probabilidad de que en al menos 2 de esos meses se hayan realizado al menos 3 instalaciones?

Solución
$Y \sim binomial(n = 6; p = P(X \geq 3) = 0.52)$
$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - 0.48^6 - 6(0.52)(0.48)^5 = 0.9083$

(4) En un centro telefónico de atención al cliente se quiere estimar el número medio de llamadas que llegan en una hora. Para ello, mediante muestreo aleatorio simple, se cuenta el número de llamadas recibidas en 50 periodos de una hora de duración, obteniéndose una media muestral igual a 35 y una desviación típica muestral igual a 6.

(a) Construye un intervalo de confianza al 90% para el número medio de llamadas por hora, especificando las hipótesis necesarias.

Solución

Sea X = "número de llamadas recibidas en una hora". Desconocemos la distribución de X así como su media y su desviación típica, que denotaremos por μ y σ respectivamente.

Mediante m.a.s. obtenemos una muestra de tamaño 50 con: $\overline{x} = 35$ y $s = 6$. Como el tamaño de muestra es suficientemente grande para poder aplicar el teorema central del límite, tenemos que el intervalo de confianza para μ es:

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp z_{\alpha/2}\, \frac{s}{\sqrt{n}}\right]$$

En nuestro caso, $n = 50$, $\alpha = 0.1$, $z_{\alpha/2} = z_{0.05} = 1.645$:

$$IC_{0.90}(\mu) = [35 \pm 1.645 \frac{6}{\sqrt{50}}] = [33.60, 36.40].$$

(b) La dirección del centro estima que el número medio de llamadas que llegan por hora es 30. Sin embargo, en los últimos meses los teleoperadores se quejan de un aumento considerable en el número de llamadas que atienden. Plantea y resuelve un contraste de hipótesis al nivel 0.05 para determinar si los teleoperadores tienen razón y establece claramente la conclusión obtenida.

Solución

El contraste a realizar sería

$$H_0 : \mu \leq 30$$
$$H_1 : \mu > 30$$

Podemos realizar el contraste de 2 formas.

1) Por el TCL sabemos que el estadístico del contraste y su distribución (aproximada) bajo la hipótesis nula son:

$$\frac{\overline{X} - 30}{\frac{S}{\sqrt{n}}} \overset{H_0}{\sim} N(0,1)$$

y por tanto la región crítica será

$$RC_{0.05} = \left\{ x_1, \ldots, x_{50} | \frac{\overline{x} - 30}{\frac{s}{\sqrt{n}}} > z_{0.05} \right\}$$

Como el valor del estadístico en la muestra es $\frac{35-30}{\frac{6}{\sqrt{50}}} = 5.89$ y $z_{0.05} = 1.645$, la muestra obtenida pertenece a la región crítica y por tanto al nivel de significación 0.05 se rechaza la hipótesis nula.

2) Calculamos el p-valor del contraste:

$$p = P\left(\frac{\overline{X}-30}{\frac{S}{\sqrt{n}}} > 5.89 \,\middle|\, H_0\right) \overset{Z\sim N(0,1)}{=} P\left(Z > 5.89\right) \approx 0.$$

Como el p-valor es menor que el nivel de significación, al nivel de significación 0.05 se rechaza la hipótesis nula.

La conclusión final es que al nivel de significación 0.05 hay suficiente evidencia estadística para rechazar la suposición de que el número medio de llamadas sea 30 por hora, es decir, se acepta que el número de llamadas ha aumentado.