Estadística I
Solución Examen Final - 28 Mayo de 2009

(1) (10 puntos) A 16 estudiantes de *Filosofía* se les preguntó cuántas clases de esta asignatura habían perdido durante el cuatrimestre. Las respuestas obtenidas fueron las siguientes: 3 1 6 1 9 1 7 2 3 0 4 4 5 0 2 0

(a) (3 puntos) Calcula la distribución de frecuencias (absolutas y relativas) de los datos anteriores, completando la siguiente tabla.

Solución

Los datos ordenados son: $0, 0, 0, 1, 1, 1, 2, 2, 3, 3, 4, 4, 5, 6, 7, 9$

Intervalo	Frecuencia absoluta	Frecuencia relativa
$[0, 2)$	6	0.3750
$[2, 4)$	4	0.2500
$[4, 6)$	3	0.1875
$[6, 8)$	2	0.1250
$[8, 10)$	1	0.0625

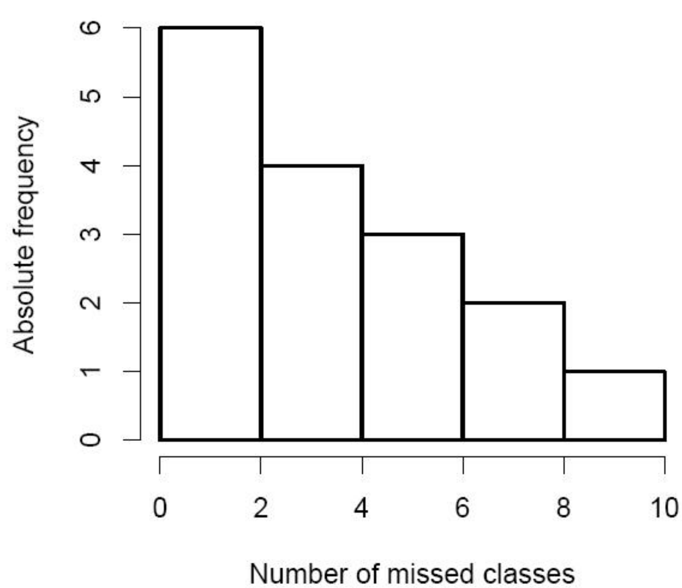
(b) (2 puntos) Observando los resultados obtenidos en el apartado (a), ¿cuántos estudiantes han perdido al menos seis clases?

Solución: 3

¿Qué porcentaje de estudiantes han perdido menos de cuatro clases?

Solución: 62.5%

(c) (2.5 puntos) Utilizando los resultados del apartado (a), dibuja el histograma de los datos (usando frecuencias absolutas) y describe la forma que tiene la distribución.



La distribución es asimétrica derecha.

(d) (1.5 puntos) Calcula la media muestral, la mediana y la moda.

$$\begin{aligned} \bar{x} &= \frac{48}{16} = 3 \\ M &= \frac{2+3}{2} = 2.5 \\ moda &= \text{ 0 y 1 (no es única)} \end{aligned}$$

(e) (1 punto) ¿Sorprende que la media sea mayor que la mediana? Justifica brevemente tu respuesta.

Solución

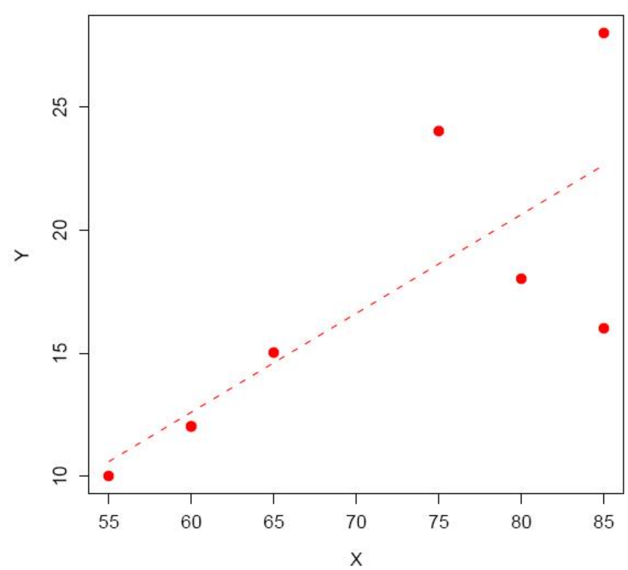
No, no estoy sorprendido porque la distribución es asimétrica derecha

(2) (10 puntos) Una compañía decide realizar un test de aptitud a los nuevos representantes de ventas. La siguiente tabla muestra el promedio de ventas semanales, Y, y la puntuación del test de aptitud (en puntos), X, para una muestra de $n = 8$ representantes.

ventas semanales promedios	10	12	28	24	18	16	15	12
puntaje	55	60	85	75	80	85	65	60

(a) (4 puntos) Dibuja el diagrama de dispersión para el promedio de ventas semanales respecto de la puntuación.

Solución



(b) (2 puntos) Se calculó la recta de regresión para el promedio de ventas mensual frente a la puntuación, usando R. Usa la salida obtenida para identificar la ordenada en el origen y la pendiente de la recta ajustada. Escribe su ecuación correspondiente.

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

Residuals:
    Min      1Q  Median      3Q     Max
-6.6514 -1.1147 -0.6009  1.6261  5.3670

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.5046     9.5745  -1.202   0.2748
X             0.4018     0.1339   3.002   0.0240 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.279 on 6 degrees of freedom
Multiple R-squared: 0.6003, Adjusted R-squared: 0.5337
F-statistic: 9.011 on 1 and 6 DF,  p-value: 0.02395
```

Solución: $\hat{Y} = -11.5046 + 0.4018 * X$

(c) (2 puntos) Dibuja la recta de regresión ajustada, en el diagrama de dispersión realizado en el apartado (a).
Ver a).

(d) (2 puntos) ¿Cuál es el valor predicho del promedio de ventas mensuales, para un representante que ha recibido 62 puntos en el test de aptitud?

Solución
$\hat{Y} = -11.5046 + 0.4018(62) = 13.407$

(3) (10 puntos) Debido a la alarma originada por la gripe de la cepa *N1H1*, el 65% de las llamadas recibidas por el teléfono nacional de emergencias están relacionadas con dudas acerca de dicha enfermedad. En un grupo de ocho llamadas elegidas al azar, halla la probabilidad de que estén relacionadas con la nueva gripe:

(a) (2 puntos) Alguna de las llamadas.
(b) (4 puntos) Más de seis.

Sea X la variable aleatoria que modeliza el número de llamadas relacionadas con la gripe, cuando se considera un total de 8 llamadas.

(c) (2 puntos) Calcula la $E[X]$.
(d) (2 puntos) Calcula la desviación típica de X.

Solución

La variable aleatoria X sigue una distribución binomial con $n = 8$, $p = 0.65$. Es decir, Bi$(8; 0,65)$.

(a)

$$P[X > 0] = 1 - P[X = 0] = 1 - 0.35^8 = 0.9998$$

(b)

$$P[X>6]=P[X=7]+P[X=8]=\begin{pmatrix}8\\7\end{pmatrix}0.65^7\cdot 0.35+\begin{pmatrix}8\\8\end{pmatrix}0.65^8=$$
$8\cdot 0.65^7\cdot 0.35+0.65^8=0.169$

(c)

$$\mu=np=8\cdot 0.65=5.2$$

(d)

$$\sigma=\sqrt{npq}=\sqrt{8\cdot 0.65\cdot 0.35}=1.35$$

(4) (10 puntos) Se sabe que la demanda diaria (en miles de unidades) de un producto fabricado por cierta empresa sigue una distribución normal. Se toma una muestra aleatoria simple de la demanda registrada en 15 días distintos, obteniéndose una media muestral igual a 3.56 y una desviación típica muestral igual a 0.92.

(a) (5 puntos) Construye un intervalo de confianza al 95% para la media de la demanda diaria de dicho producto, especificando las hipótesis necesarias.

Solución
Sea X = "demanda diaria del producto (en miles de unidades)". Se sabe que $X\sim N(\mu,\sigma)$.
Mediante m.a.s. obtenemos una muestra de tamaño 15 con: $\overline{x}=3.56$ y $s=0.92$.
Intervalo de confianza para μ en una población normal con varianza desconocida:

$$IC_{1-\alpha}(\mu)=[\bar{x}\pm t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}]$$

En nuestro caso, $n=15$, $\alpha=0.05$, $t_{n-1;\alpha/2}=t_{14;0.025}=2.145$:

$$IC_{0.95}(\mu)=[3.56\pm 2.145\frac{0.92}{\sqrt{15}}]=[3.05,4.07].$$

(b) (5 puntos) El plan de producción de la empresa está basado en la suposición de que la demanda media diaria del citado producto es de 4000 unidades. Contrasta esta hipótesis, al nivel de significación de 0.05, e indica claramente la conclusión obtenida. Indica quienes son las hipótesis nula y alternativa.

Solución
El contraste a realizar sería

$$H_0:\mu=4$$
$$H_1:\mu\neq 4$$

Podemos realizar el contraste de 3 formas.
1) Como el contraste es bilateral y $4\in[3.05,4.07]=IC_{0.95}(\mu)$, por la equivalencia entre contrastes de hipóteses e intervalos de confianza, sabemos que al nivel de significación 0.05 no hay suficiente evidencia

muestral para rechazar la hipótesis nula.
2) Bajo las hipótesis de normalidad y varianza desconocida, el estadístico del contraste es

$$\frac{\overline{X}-4}{\frac{S}{\sqrt{n}}} \stackrel{H_0}{\sim} t_{n-1}$$

y por tanto la región crítica será

$$RC_{0.05} = \left\{ x_1,\ldots,x_{15} \middle| \left| \frac{\overline{x}-4}{\frac{s}{\sqrt{n}}} \right| > t_{14;0.025} \right\}$$

Como el valor del estadístico en la muestra es $\frac{3.56-4}{\frac{0.92}{\sqrt{15}}} = -1.85$ y $t_{14;0.025} = 2.145$, la muestra obtenida no pertenece a la región crítica y por tanto al nivel de significación 0.05 no se rechaza la hipótesis nula.
3) Calculamos el p-valor del contraste:

$$p = P\left(\left| \frac{\overline{X}-4}{\frac{S}{\sqrt{n}}} \right| > |-1.85| \middle| H_0 \right) = P\left(|t_{14}| > 1.85\right) = 2\cdot P(t_{14} > 1.85) \in (0.05, 0.1).$$

Como el p-valor es mayor que el nivel de significación, al nivel de significación 0.05 no se rechaza la hipótesis nula.

La conclusión final es que al nivel de significación 0.05 no hay suficiente evidencia estadística para rechazar la suposición de que la demanda media diaria sea igual a 4.