Parcial 2 Estadística

Juan Sebastian Velasquez Acevedo
1744936-3743
juan.velasquez.acevedo@correounivalle.edu.co

Índice

Punto 1 **2**
- 1.a 2
- 1.b 3
- 1.c 4
- 1.d 5
- 1.e 6

Punto 2 **7**
- 2.a - 2.b - 2.c - 2.d 7
- 2.e 7

Punto 3 **9**
- 3.a 9
- 3.b 10
- 3.c 11

Punto 1

1.a

Parcial 2 - estadística

Juan Sebastian Velasquez Acevedo

1744936

1) $$f(x) = \begin{cases} k(3-x^2) & , -1 \leq x \leq 1 \\ 0 & , \text{en otro caso} \end{cases}$$

a) Por propiedad : $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\rightarrow \cancel{\int_{\infty}^{-1} 0 \, dx} + \int_{-1}^{1} k(3-x^2)\,dx + \cancel{\int_{1}^{\infty} 0 \, dx} = 1$$

$$\rightarrow \int_{-1}^{1} K(3-x^2)\,dx = K \int_{-1}^{1} (3-x^2)\,dx$$

$$\rightarrow K\left(3x - \frac{x^3}{3}\right)\Big|_{-1}^{1} = K\left(\left(3-\frac{1}{3}\right) - \left(-3+\frac{1}{3}\right)\right)$$

$$\rightarrow K\left(3 - \frac{1}{3} + 3 - \frac{1}{3}\right) = K\left(\frac{8}{3} + \frac{8}{3}\right) = K\left(\frac{16}{3}\right)$$

Como: $\int_{-1}^{1} K(3-x^2)\,dx = 1$

$$K\left(\frac{16}{3}\right) = 1$$

$$K = \frac{3}{16}$$

1.b

1.b) Probabilidad con error aleatorio 'x' en la medición:

Variable aleatoria continúa

La función de distribución acumulada para una variable aleatoria continúa se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

$\rightarrow$ $P(X < 1/2)$ para $-1 \leq x \leq 1$ según función densidad.

La función acumulada (fda):

a) Para cualquier menor a -1:

$$F(-\infty) = 0$$

b) Para cualquier mayor a 1:

$$F(\infty) = 1$$

c) Para cualquier número entre -1 y 1:

$$F(x) = \int_{-1}^{x} 3/16\,(3 - t^2)\,dt = \frac{3}{16}\left(3t - \frac{t^3}{3}\right)\Big|_{-1}^{x}$$

$$F(x) = \frac{3}{16}\left(\left(3x - \frac{x^3}{3}\right) - \left(-3 + \frac{1}{3}\right)\right)$$

$$F(x) = \frac{3}{16}\left(3x + 3 - \frac{x^3 + 1}{3}\right) = \frac{-x^3 + 9x + 8}{16}$$

$$F(x) = \frac{-x^3 + 9x + 8}{16}$$

$\longrightarrow \quad F(x) = \dfrac{-x^3 + 9x + 8}{16}$

Función de distribución acumulada:

$$F(x) = \begin{cases} 0 & , \ x < -1 \\ \dfrac{-x^3+9x+8}{16} & , \ -1 \leq x \leq 1 \\ 1 & , \ x > 1 \end{cases}$$

$$P(X < 1/2) = F(1/2) = \frac{(-1/2)^3 + 9(1/2) + 8}{16}$$

$$= \frac{(99/8)}{16} = \frac{99}{(8)(16)} \approx \frac{99}{128}$$

1.c

c) $P(|X| > 0{,}8)$

Propiedad de valor absoluto:

$$P(X < -0.8) + P(X > 0.8) = F(-0.8) + (1 - F(0.8))$$

$$= 0{,}082 + (1 - 0.918) = 0.082 + 0.082$$

$$= 0.164$$

$$F(-0.8) = \frac{(0.8)^3 + 9(-0.8) + 8}{16} = 0.082$$

$$F(0.8) = \frac{(-0.8)^3 + 9(0.8) + 8}{16} = 0.918$$

1.d

d) Momentos de la variable aleatoria: $E(x)$ y $V(x)$

El <u>valor esperado</u> de una variable aleatoria X es el valor medio de X, se define como $E(X)$ o μ_X y está dado por:

$$E(x) = \int_{-\infty}^{\infty} x \underbrace{f(x)}_{\text{función de densidad de probabilidad}} dx \quad \text{si } X \text{ es continúa}$$

Tenemos que: $f(x) = \frac{3}{16}(3 - x^2)$

$$\rightarrow E(x) = \int_{-1}^{1} x\left(\frac{3}{16}(3 - x^2)\right) dx = \frac{3}{16}\int_{-1}^{1} 3x - x^3\, dx \rightarrow$$

$$\rightarrow \frac{3}{16}\left(\frac{3x^2}{2} - \frac{x^4}{4}\right)\Big|_{-1}^{1} = \frac{3}{16}\left(\left(\frac{3}{2} - \frac{1}{4}\right) - \left(\frac{3}{2} - \frac{1}{4}\right)\right)$$

$$\rightarrow 0 \qquad E(x) = 0$$

La <u>varianza</u> de una variable aleatoria X está definida como:

$$Var(x) = \sigma^2 = E(x^2) - (E(x))^2$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$E(x^2) = \int_{-1}^{1} x^2 \frac{3}{16}(3 - x^2)\, dx = \frac{3}{16}\int_{-1}^{1} 3x^2 - x^4\, dx$$

$$= \frac{3}{16}\left(x^3 - \frac{x^5}{5}\right)\Big|_{-1}^{1} = \frac{3}{16}\left(\left(1 - \frac{1}{5}\right) - \left(-1 + \frac{1}{5}\right)\right)$$

$$= 3/16\,(4/5 + 1 - 1/5) = 3/16\,(1 + 3/5) = \frac{3}{16}\left(\frac{8}{5}\right)$$

$$= \frac{24}{80} = 0,3$$

$$\rightarrow \underline{V(x)} = E(x^2) - (E(x))^2 = 0,3 - (0)^2 = 0,3$$

1.e

e) Percentil 80 de la distribución.

Definición de percentil de una distribución continúa:

Sea X una V.a continua con función de densidad $f(x)$ y función de distribución acúmulada $F(x)$ y sea: $0 < p < 1$, el percentil $(100\,p)$-ésimo de la distribución de X es el valor x_p tal que:

$$F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(t)\,dt = p$$

$$\rightarrow F(x_p) = P(X \leq x_p) = 0.8$$

Tenemos nuestra función de distribución acúmulada:

$$F(x) = \frac{-x^3 + 9x + 8}{16} \quad \text{cuando} \quad \underline{\underline{-1 \leq x \leq 1}}$$

$$\rightarrow F(x_p) = \frac{-x_p^3 + 9x_p + 8}{16} = 0.8$$

$$\rightarrow F(x_p) = -5x_p^3 + 45x_p - 24 = 0$$

3 raíces: $x_p \approx 0.55202$, $x_p \approx 2.68565$, $x_p \approx -3.23767$

tal que $-1 \leq x \leq 1$, se escogé la raíz en este rango:

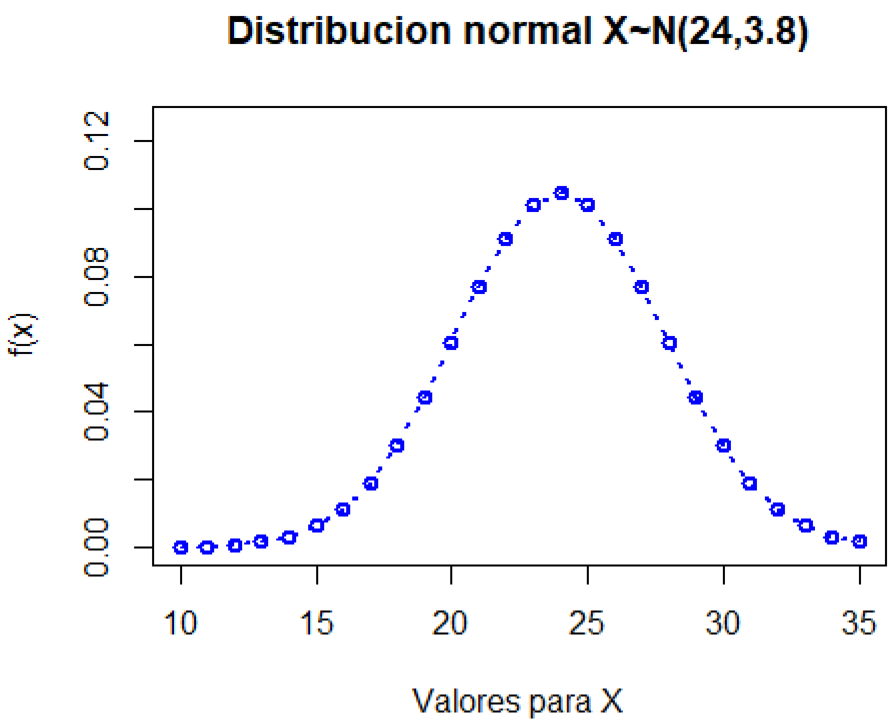
$$\boxed{x_p \approx 0.55202}$$

Punto 2

2.a - 2.b - 2.c - 2.d

Los subpuntos **2.a, 2.b, 2.c, 2.d** están desarrollados y explicados en el siguiente repositorio:
https://github.com/Odzen/Probability_Statistics/blob/main/Parciales/parcial2/r_script_parcial_2.R

Gráfico de distribución normal:



2.e

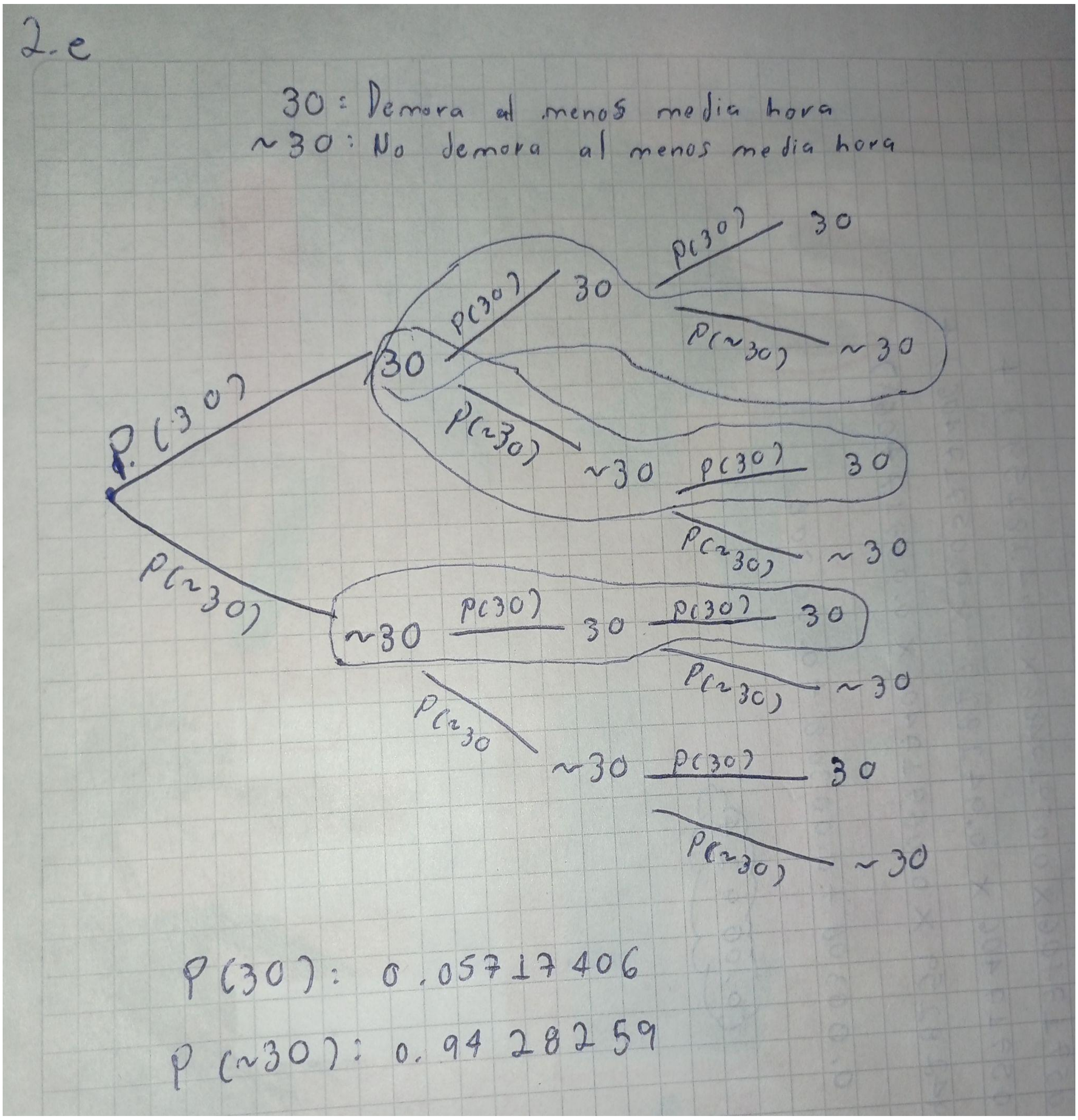
Para este punto tomamos los valores calculados en el punto 2.a:

P(X>30) = 0.05717406

A partir de lo anterior podemos calcular:

P(X<=30) = 1 - P(X>30) = 0.9428259

Realizamos árbol de probabilidades:



2.e: A partir del árbol se cálcula X.

Siendo X = Probabilidad de que 2 de los siguientes 3 viajes tomen al menos 1/2 hora.

$$X = 2 \rightarrow (P(30) \times P(30) \times P(\sim 30)) +$$
$$(P(30) \times P(\sim 30) \times P(30)) +$$
$$(P(\sim 30) \times P(30) \times P(30)) +$$

$$\rightarrow (0.05717406 \times 0.05717406 \times 0.9428259) +$$
$$(0,05717406 \times 0,9428259 \times 0,05717406) +$$
$$(0,9428259 \times 0.05717406 \times 0.05717406)$$

$$\rightarrow 0,00308 + 0,00308 + 0,00308$$

$$X \rightarrow \boxed{0,00924}$$

Punto 3

3.a

¿Qué puede concluir sobre la relación entre la variable respuesta y las variables regresoras?

Variable de respuesta:

Y: Precio de venta de la propiedad

Variables regresoras:

X1: Impuestos que paga anualmente

X8: Edad de la propiedad

Según el modelo presentado se puede concluir que

- Entre la variable regresora X1 y la variable de respuesta Y hay una relación lineal directa moderada. Esto se ve reflejado en el coeficiente de correlación de Pearson, se aprecia que tiene un valor de 0.8759762, un valor positivo cercano a 1, lo que indica que los puntos efectivamente tienen una tendencia a disponerse alineadamente.
- Entre la variable regresora X8 y la variable de respuesta Y hay una relación lineal inversa débil. Esto se ve reflejado en el coeficiente de correlación de Pearson, se aprecia que tiene un valor de -0.3985177, un valor negativo. El valor negativo nos indica que los puntos tienen una tendencia inversa y también al estar un poco lejos de -1, podemos decir que hay una tendencia débil de que los puntos se dispongan alineadamente.

3.b

Ecuación lineal del modelo:

Coeficientes
$\beta_0 = 13.834472$ → Intercepto con el eje Y. Es la ordenada en el origen, el valor de la variable dependiente Y cuando todos los predictores son cero.
$\beta_1 = 3.292349$ → es el efecto promedio que tiene el incremento en una unidad de la variable predictora X1 sobre la variable dependiente Y, manteniéndose constantes el resto de variables.
$\beta_2 = -0.008254$ → es el efecto promedio que tiene el incremento en una unidad de la variable predictora X2 sobre la variable dependiente Y, manteniéndose constantes el resto de variables.
ε → Error aleatorio

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$Y = 13.834472 + 3.292349\, X_1 - 0.008254\, X_2 + \varepsilon$$

Los coeficientes los podríamos interpretar de la siguiente forma:

β_0 → El precio de venta promedio de la propiedad se espera que sea aproximadamente 13.8 millones cuando los impuestos que se pagan anualmente y la edad de la propiedad son 0.

β_1 → Significa que por cada millón de pesos adicional que paga de impuestos anualmente, existiría un aumento de aproximadamente 3.3 millones en el precio de venta de la propiedad.

β2 → Significa que por cada año que pasa desde que se construyó la propiedad, el precio de venta de la propiedad disminuiría en un cantidad aproximada de 0.0082 millones.

3.c

Sabemos que la proporción de variabilidad de Y atribuido a la relación entre X1, X2 e Y; está dado por el coeficiente de determinación (r^2) el cual determina la calidad del modelo para replicar los resultados.

Coeficiente de determinación es igual al coeficiente de correlación al cuadrado = r^2 y este valor siempre está entre 0 y 1

Se sabe que cuanto más cerca esté r^2 de 1 mejor será el ajuste del modelo lineal y entre más cerca esté r^2 de 0 peor será el ajuste del modelo lineal.

Según los datos dados tenemos que el coeficiente de determinación es igual a 0.7676, está bastante cerca de 1, lo que indica que gran parte de la variación de los valores observados de Y puede ser explicados por el modelo de regresión lineal múltiple construido.

Tener r^2 = 0.7676 indica también que el 76.76% de la variabilidad del precio de venta de la propiedad pudo ser explicada por el modelo propuesto.