

Control 1

MAT041
1er Semestre 2022

Pregunta 1 (60 pts)

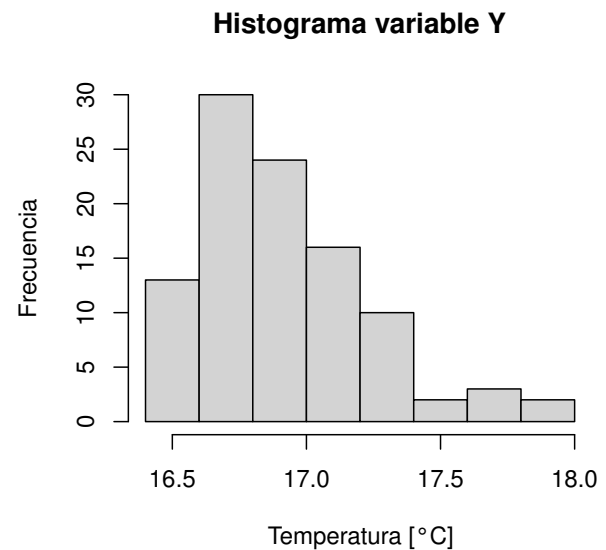
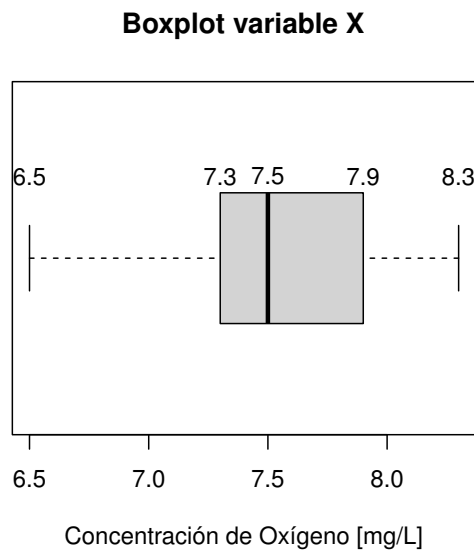
Un grupo de científicos se encuentra realizando un estudio sobre las condiciones de vida marina en una localidad de Chile; para esto han tomado 100 muestras de agua de mar al azar en distintos puntos de la costa de Valparaíso, estas muestras se han almacenado en probetas a las cuales se les han medido las siguientes variables:

X = Concentración de oxígeno disuelto [mg/L].

Y = Temperatura al momento de la medición [$^{\circ}C$].

Z = Presencia de materiales sólidos (medida como **si** o **no**).

La información obtenida de las variables X e Y se ha resumido en las siguientes gráficas:



1. Describa el tipo de muestreo que se llevó a cabo y caracterice los tipos de variables medidas junto con su escala de medición. [15 puntos]
2. Para la variable X describa lo observado en el Boxplot, luego reporte una medida de tendencia central y calcule una medida de dispersión [25 puntos]
3. Describa el histograma de la variable Y , proponga medidas de tendencia central y de variabilidad adecuadas para ser calculadas [20 puntos]

Solución:

1. La muestra se obtuvo a través de un muestreo aleatorio simple. La variable X es de tipo cuantitativa (o continua) (y/o acotada) medida en escala de razón, la variable Y es de tipo cuantitativa (continua y no acotada) medida en escala intervalar y la variable Z es de tipo discreta medida en escala nominal.

2. En el gráfico de boxplot se observa que la mediana está más cerca del cuartil 1 que del cuartil 3, esto podría indicar asimetría positiva. No se observan valores escapados (o outliers). De la información que se puede extraer del boxplot podemos obtener como medida de tendencia central la mediana $Me = 7.5$ y como medida de variabilidad el rango intercuartil $RIQ = 7.9 - 7.3 = 0.6$
3. Se observa que el histograma de la variable Y presenta una asimetría positiva (o sesgo positivo) y valores que eventualmente podrían ser datos escapados (o outliers). Debido a esto se recomienda utilizar medidas de tendencia central robustas como la mediana o la moda, para el caso de la variabilidad se recomiendan medidas robustas como el rango intercuartil.

Pregunta 2 (40 pts)

Una método que se utiliza para estimar el volumen de los pinos es considerarlos como conos y realizar el cálculo mediante $V = \frac{1}{3}\pi r^2 \cdot h$, Para verificar esto se han tomado 31 mediciones de pinos en los cuales se han medido el volumen real (Y) y además el volumen aproximado (X) mediante la fórmula anterior. De este experimento se reporta a siguiente información:

$$\bar{X} = 26, \quad \bar{Y} = 30$$

$$S_X = 13 \quad S_Y = 16, \quad \text{cov}(X, Y) = 206$$

1. Encuentre e interprete el valor de la correlación entre X e Y . ¿Cree usted que este método es una buena forma de estimar el valor del volumen de un pino? [20 puntos]
2. Encuentre los valores estimados de los parámetros β_0 y β_1 en el modelo lineal $Y = \beta_0 + \beta_1 X$. [20 puntos]

Solución

- Usando directamente la fórmula para la correlación

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{S_x \cdot S_y} = \frac{206}{13 \cdot 16} \approx 0.99,$$

este valor denota una correlación lineal directa entre las variables X e Y bastante buena, por lo que se puede interpretar que el método es una buena forma de estimar el volumen real del pino.

- Usando directamente las ecuaciones para $\hat{\beta}_0$ y $\hat{\beta}_1$ se obtiene

$$\hat{\beta}_1 = \frac{\text{Cov}(X, Y)}{S_X^2} = \frac{206}{13^2} \approx 1.22$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \approx -1.72$$