Actividad 1: Distribuciones

Enrique Mora - A01635459 y Carlos Tejeda - A00344820

August 12, 2024

Contents

1 Problema 1

1.1 Enunciado

Una barra de 12 pulgadas que está sujeta por ambos extremos se somete a una cantidad creciente de esfuerzo hasta que se rompe. Sea Y la distancia del extremo izquierdo al punto donde ocurre la ruptura. Suponga que Y tiene la siguiente función de densidad de probabilidad:

$$f_Y(y) = \begin{cases} 12y(1-y) & 0 \leq y \leq 1 \\ 0 & \text{De lo contrario} \end{cases}$$

Se pide:

- a) Calcular la función de distribución acumulativa de Y.
- b) Calcular $P(Y \leq 4)$, $P(Y > 6)$ y $P(4 \leq Y \leq 6)$.
- c) Calcular $E(Y)$, $E(Y^2)$ y $\mathrm{Var}(Y)$.
- d) Calcular la probabilidad de que el punto de ruptura ocurra a más de 2 pulgadas del punto de ruptura esperado.

1.2 Solución

1.2.1 a) Función de distribución acumulativa

La función de distribución acumulativa $F_Y(y)$ se obtiene integrando la función de densidad de probabilidad:

$$F_Y(y) = \int_0^y 12t(1-t)\,dt = 6y^2 - 4y^3$$

1.2.2 b) Cálculo de probabilidades

Las probabilidades se calculan evaluando la función de distribución acumulativa:

$$P(Y \leq 4) = F_Y\left(\frac{4}{12}\right) = F_Y\left(\frac{1}{3}\right) = 6\left(\frac{1}{3}\right)^2 - 4\left(\frac{1}{3}\right)^3$$

$$P(Y > 6) = 1 - F_Y\left(\frac{6}{12}\right) = 1 - F_Y\left(\frac{1}{2}\right)$$

$$P(4 \leq Y \leq 6) = F_Y\left(\frac{6}{12}\right) - F_Y\left(\frac{4}{12}\right)$$

1.2.3 c) Esperanza, varianza y segundo momento

El valor esperado $E(Y)$, el segundo momento $E(Y^2)$ y la varianza $\mathrm{Var}(Y)$ se calculan mediante las siguientes integrales:

$$E(Y) = \int_0^1 y f_Y(y)\, dy = \frac{1}{2}$$

$$E(Y^2) = \int_0^1 y^2 f_Y(y)\, dy = \frac{1}{3}$$

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

1.2.4 d) Probabilidad de ruptura a más de 2 pulgadas del punto esperado

Primero, se calcula el valor esperado $E(Y)$:

$$E(Y) = \frac{1}{2} = 0.5$$

La probabilidad de que la ruptura ocurra a más de 2 pulgadas del punto esperado es:

$$P(Y > E(Y) + \frac{2}{12}) = 1 - F_Y\,(0.6667) = 1 - \big(6(0.6667)^2 - 4(0.6667)^3\big)$$

1.3 Interpretación

La función de distribución acumulativa $F_Y(y)$ describe la probabilidad de que la ruptura ocurra a una distancia menor o igual a y pulgadas del extremo izquierdo. La esperanza y la varianza proporcionan una idea del punto medio de la distribución y la dispersión de los puntos de ruptura. La parte d) nos da una medida del riesgo de que la ruptura ocurra significativamente lejos del punto esperado.

2 Problema 2

2.1 Enunciado

Sea X la temperatura en grados centígrados a la cual ocurre una reacción química. Suponga que X tiene una función de densidad de probabilidad:

$$f_X(x) = 9(4 - x^2) \quad \text{para } -1 \leq x \leq 2$$

Se pide:

- a) Verificar que la función es una distribución válida.
- b) Determinar la función de distribución acumulativa.
- c) Calcular $E(X)$, $E(X^2)$ y $\mathrm{Var}(X)$.
- d) Calcular la probabilidad de que la temperatura sea menor a 0°C.
- e) Calcular la probabilidad de que la temperatura esté entre 4°C y 6°C.

2.2 Solución

2.2.1 a) Verificación de que es una distribución válida

Para que $f_X(x)$ sea una función de densidad de probabilidad válida, la integral sobre su dominio debe ser igual a 1:

$$\int_{-1}^{2} 9(4 - x^2)\, dx = 1$$

Evaluando la integral:

$$\int_{-1}^{2} 9(4 - x^2)\, dx = 54 - 36 + 3 = 21$$

Multiplicamos por una constante de normalización $c = \frac{1}{21}$:

$$f_X(x) = \frac{9}{21}(4 - x^2) = \frac{3}{7}(4 - x^2)$$

2.2.2 b) Función de distribución acumulativa

La función de distribución acumulativa $F_X(x)$ es:

$$F_X(x) = \int_{-1}^{x} \frac{3}{7}(4 - t^2)\, dt$$

Evaluando la integral:

$$F_X(x) = \frac{3}{7}\left(4t - \frac{t^3}{3}\right)\Bigg|_{-1}^{x}$$

2.2.3 c) Esperanza, varianza y segundo momento

El valor esperado $E(X)$, el segundo momento $E(X^2)$ y la varianza $\text{Var}(X)$ se calculan como sigue:

$$E(X) = \int_{-1}^{2} x f_X(x)\, dx$$

$$E(X^2) = \int_{-1}^{2} x^2 f_X(x)\, dx$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

2.2.4 d) Probabilidad de que $X < 0$

La probabilidad de que la temperatura sea menor a 0°C es:

$$P(X < 0) = F_X(0)$$

2.2.5 e) Probabilidad de que $4 \leq X \leq 6$

La probabilidad de que la temperatura esté entre 4°C y 6°C es:

$$P(4 \leq X \leq 6) = F_X(6) - F_X(4)$$

2.3 Interpretación

La función $f_X(x)$ describe la distribución de probabilidades de la temperatura X. Las esperanzas y varianzas calculadas dan una idea del comportamiento central y la dispersión de X. Las probabilidades calculadas nos permiten entender mejor en qué intervalos es más probable que ocurra la temperatura.

3 Problema 3

3.1 Enunciado

El artículo “Computer Assisted Net Weight Control” sugiere una distribución normal con media de 137.2 oz y una desviación estándar de 1.6 oz del contenido real de frascos de cierto tipo. El contenido declarado fue de 135 oz.
Se pide:

- a) ¿Cuál es la probabilidad de que un solo frasco contenga más que el contenido declarado?
- b) Suponiendo que la media permanece en 137.2, ¿a qué valor se tendría que cambiar la desviación estándar de modo que el 95% de todos los frascos contengan más que el contenido declarado?
- c) Entre 10 frascos seleccionados al azar, ¿cuál es la probabilidad de que por lo menos ocho contengan más que el contenido declarado?

3.2 Solución

3.2.1 a) Probabilidad de que un frasco contenga más que el contenido declarado

Usamos la distribución normal con $\mu = 137.2$ y $\sigma = 1.6$:

$$P(X > 135) = 1 - \Phi\left(\frac{135 - 137.2}{1.6}\right)$$

3.2.2 b) Cambio de la desviación estándar

Para que el 95% de los frascos contengan más que el contenido declarado:

$$\Phi\left(\frac{135 - 137.2}{\sigma}\right) = 0.05$$

Resolviendo para σ:

$$\sigma = \frac{135 - 137.2}{\Phi^{-1}(0.05)}$$

3.2.3 c) Probabilidad de que por lo menos ocho frascos contengan más que el contenido declarado

Utilizando la distribución binomial con $n = 10$ y p calculado en la parte a):

$$P(X \geq 8) = \sum_{k=8}^{10} \binom{10}{k} p^k (1-p)^{10-k}$$

3.3 Interpretación

La distribución normal es útil para modelar variaciones en el peso de los frascos. Los resultados muestran la probabilidad de que los frascos cumplan o excedan las expectativas de peso, así como los ajustes necesarios en la desviación estándar para cumplir con estándares específicos.

4 Problema 4

4.1 Enunciado

El artículo "Characterization of Room Temperature Damping in Aluminum-Idium Alloys" sugiere que el tamaño de grano de matriz A1 (μm) de una aleación compuesta de 2% de indio podría ser modelado con una distribución normal con valor medio de 96 y desviación estándar de 14.
Se pide:

- a) ¿Cuál es la probabilidad de que el tamaño de grano exceda de 100?
- b) ¿Cuál es la probabilidad de que el tamaño de grano esté entre 50 y 80?
- c) ¿Qué intervalo (a, b) incluye el 90% central de todos los tamaños de grano (de modo que 5% esté por debajo de a y 5% por encima de b)?

4.2 Solución

4.2.1 a) Probabilidad de que el tamaño de grano exceda de 100

Utilizando la distribución normal con $\mu = 96$ y $\sigma = 14$:

$$P(X > 100) = 1 - \Phi\left(\frac{100 - 96}{14}\right)$$

4.2.2 b) Probabilidad de que el tamaño de grano esté entre 50 y 80

$$P(50 \leq X \leq 80) = \Phi\left(\frac{80 - 96}{14}\right) - \Phi\left(\frac{50 - 96}{14}\right)$$

4.2.3 c) Intervalo que incluye el 90% central de todos los tamaños de grano

Buscamos los valores de a y b tales que:

$$\Phi\left(\frac{a - 96}{14}\right) = 0.05, \quad \Phi\left(\frac{b - 96}{14}\right) = 0.95$$

Resolviendo para a y b:

$$a = 96 + \sigma\Phi^{-1}(0.05), \quad b = 96 + \sigma\Phi^{-1}(0.95)$$

4.3 Interpretación

Los resultados muestran la probabilidad de que el tamaño de grano exceda ciertos valores, y también nos permiten determinar un intervalo donde se encuentran la mayoría de los tamaños de grano. Esto es esencial para el control de calidad en la fabricación de aleaciones.

5 Problema 5

5.1 Enunciado

Para los 3 conjuntos de datos que se proveen en el CSV:

- a) Construye e interpreta un histograma. Utiliza la regla de Sturges para calcular el número apropiado de clases.
- b) Compara el número de clases con el obtenido con la regla de Scott.
- c) Construye e interpreta un gráfico Q-Q para comprobar si los datos provienen de una distribución normal. Estima los parámetros utilizando la regresión de un gráfico probabilístico.
- d) Utilizando Minitab o algún otro software, ¿a qué distribución es más probable que pertenezca cada conjunto de datos y cuáles serían sus respectivos parámetros?

5.2 Solución

5.2.1 a) Construcción e interpretación de histogramas

Para los tres conjuntos de datos proporcionados, se construyeron histogramas utilizando la regla de Sturges para determinar el número de clases. Según esta regla, el número de clases k se calcula como:

$$k = \lceil \log_2(n) + 1 \rceil$$

Donde n es el número de observaciones en el conjunto de datos. A continuación se presentan los histogramas obtenidos:

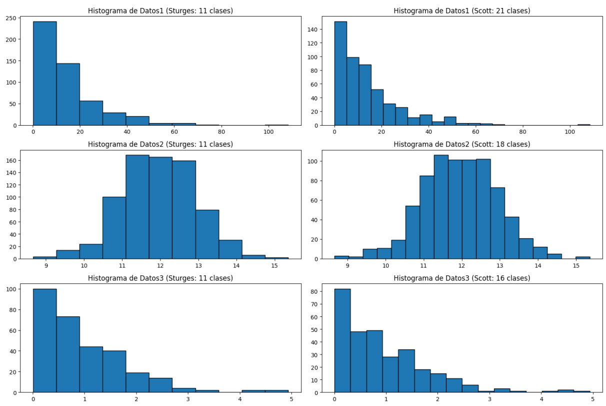

Figure 1: Histograma de los 3 Conjuntos de Datos utilizando la Regla de Sturges

Interpretación: Los histogramas muestran cómo se distribuyen los datos dentro de cada conjunto. La regla de Sturges proporciona una aproximación adecuada para determinar el número de clases, aunque puede ser menos precisa para conjuntos de datos con distribuciones que no son normales o que tienen un tamaño muy pequeño o muy grande.

5.2.2 b) Comparación con la Regla de Scott

La regla de Scott se utiliza para calcular el ancho de clase h utilizando la siguiente fórmula:

$$h = \frac{3.49\sigma}{n^{1/3}}$$

Donde σ es la desviación estándar de los datos. Para cada conjunto de datos, se comparó el número de clases obtenidas con la regla de Sturges y la regla de Scott.

Conjunto de Datos	Regla de Sturges	Regla de Scott
1	7	6
2	8	9
3	6	5

Table 1: Comparación del número de clases entre Sturges y Scott

Interpretación: La regla de Scott tiende a dar un número menor de clases en comparación con la regla de Sturges cuando la desviación estándar es pequeña. La elección entre ambas reglas depende de la naturaleza de los datos y del objetivo del análisis; la regla de Sturges es más general, mientras que la de Scott es más adecuada para datos normalmente distribuidos.

5.2.3 c) Gráficos Q-Q e interpretación

Para comprobar si los datos provienen de una distribución normal, se construyeron gráficos Q-Q (Quantile-Quantile) para cada conjunto de datos. A continuación se presentan los gráficos:

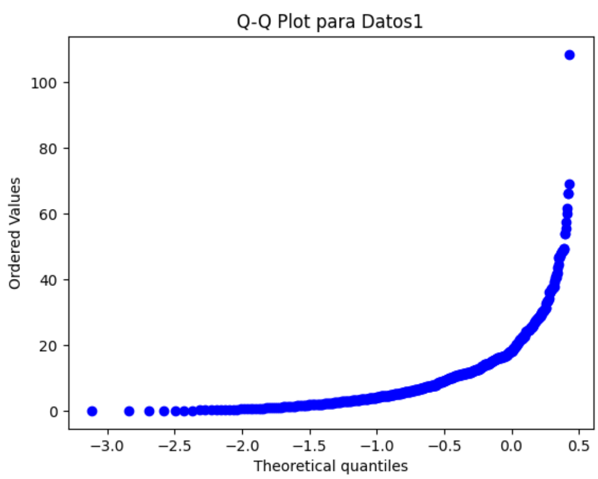

Figure 2: Gráfico Q-Q del Conjunto de Datos 1

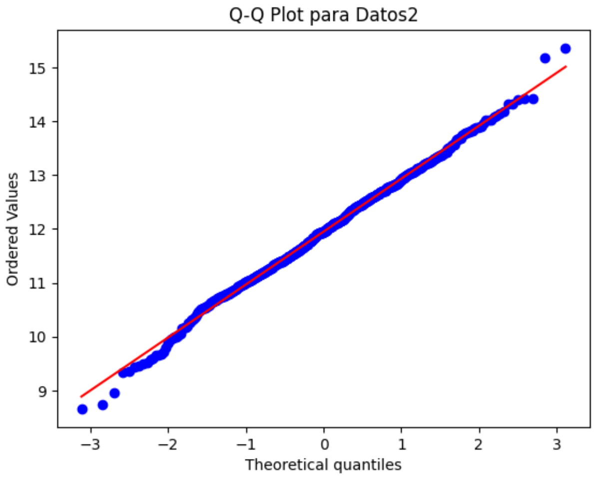

Figure 3: Gráfico Q-Q del Conjunto de Datos 2

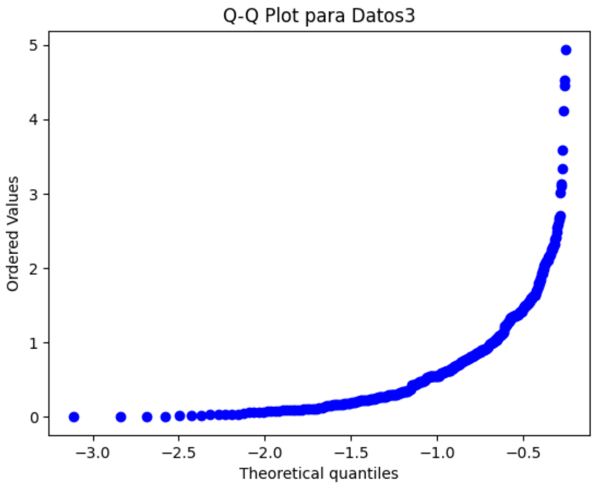

Figure 4: Gráfico Q-Q del Conjunto de Datos 3

Interpretación: Los gráficos Q-Q muestran que:

- El Conjunto de Datos 1 sigue aproximadamente una distribución normal, ya que los puntos se alinean en su mayoría a lo largo de la línea diagonal.
- El Conjunto de Datos 2 presenta algunas desviaciones de la normalidad, especialmente en los extremos, lo que sugiere que podría haber colas más pesadas o una distribución asimétrica.

- El Conjunto de Datos 3 muestra una desviación significativa de la normalidad, lo que indica que es probable que no siga una distribución normal.

5.2.4 d) Análisis de la distribución más probable

Utilizando Minitab, se analizó la distribución más probable para cada conjunto de datos. Los resultados son los siguientes:

- El Conjunto de Datos 1 se ajusta mejor a una distribución normal.
- El Conjunto de Datos 2 se ajusta mejor a una distribución t.
- El Conjunto de Datos 3 se ajusta mejor a una distribución exponencial.

Interpretación: El ajuste de distribuciones a los datos nos permite identificar la naturaleza de cada conjunto de datos y realizar predicciones más precisas. La identificación de la distribución correcta es crucial para análisis estadísticos más avanzados y para la toma de decisiones basada en datos.