Probabilidad y estadística

Gabriela María Castro Beltrán, Juan Nicolás Carvajal Useche *gcastrob@unal.edu.co*, *jcarvajalu@unal.edu.co*
Universidad Nacional de Colombia

I. Cuestionario

1. Un elemento en voladizo se carga y la deflexión es medida en la punta. Los siguientes resultados se obtuvieron con 10 repeticiones de carga (deflexión en cm)

5.45	5.73	6.77	6.23	5.43	5.45	6.09	5.66	5.82	5.71

Figura 1. Deflexión del elemento en voladizo

Determine el valor medio, la desviación estándar y la mejor estimación del valor real de este conjunto de datos. En qué intervalo deberá estar el 50% de toda la población de los valores de deflexión?

Para calcular el valor medio (media aritmética) de este conjunto de datos, debemos sumar todos los valores x y dividirlos entre el número total de valores N:

$$\frac{\sum_{i=0}^{N} x_i}{N} = \frac{58,34}{10} = 5,834cm$$

Para calcular la desviación estándar, primero debemos calcular la varianza, que es la media de las diferencias cuadráticas entre cada valor y la media (μ):

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2 = 0,179cm$$

La desviación estándar es la raíz cuadrada de la varianza:

$$\sigma = 0,423cm$$

Para la mejor estimación del valor real de este conjunto de datos es la media que es igual a 5,834 cm.

Para determinar el intervalo en el que se encuentra el 50% de toda la población de los valores de deflexión, necesitamos calcular la mediana, la cual es el valor que se encuentra en el centro del conjunto de datos cuando se ordenan en orden creciente. Para este caso :

$$M_e = (5,73 + 5,71)/2 = 5,72$$

Para encontrar el intervalo, podemos buscar los valores que se encuentran a 25% y 75% de los datos ordenados. Estos vienen dados por los cuartiles Q1 y Q3, respectivamente. Para encontrar Q1, buscamos el valor que se encuentra en la posición (n+1)/4 y para encontrar Q3, buscamos el valor que se encuentra en la posición 3(n+1)/4, donde n es el número de datos.

$$Q_1 = (10+1)/4 = \text{Posición } 2,75 = 5,45$$

$$Q_3 = 3(10+1)/4 = \text{Posición } 8,25 = 6,125$$

El intervalo que contiene el 50% de los datos es [Q1, Q3], que es [5,45, 6,125].

2. Resuelva los siguientes problemas haciendo uso de la distribución normal:

Sugerencia: Use las tablas de valores para la distribución normal, es posible que tenga que normalizar primero usando la media y la desviación.

1. Si X es una variable aleatoria de una distribución $N(\mu,\sigma)$, hallar: $P(\mu-3\sigma \leq X \leq \mu+3\sigma)$

   Para calcular la probabilidad se utiliza la tabla z, en donde es necesario tener una distribución normal estándar, en donde:

$$z_1 = \frac{x-\mu}{\sigma} = \frac{(\mu+3\sigma)-\mu}{\sigma} = 3$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{(\mu-3\sigma)-\mu}{\sigma} = -3$$

   Se hace uso de la tabla de la Distribución Normal Estándar N(0,1), en donde para una z =3, tiene una probabilidad de 0.0013500 y para una z=-3 la probabilidad es igual a 1-0.0013500=0,99865. De esta manera la probabilidad de que x se encuentre entre $\mu - 3\sigma$ y $\mu + 3\sigma$ es igual a 0,99865-0.0013500= 0,9973

2. En una distribución normal de media 4 y desviación estándar 1,5, calcular el valor de a para que: $P(4-a \leq X \leq 4+a) = 0,5934$

   Se estandariza la distribución normal para calcular la probabilidad. En este caso, se tiene:

$$z = \frac{(x-\mu)}{\sigma}$$

$$0,5934 = \frac{x-4}{1,5}$$

$$x = 4,8901$$

$$a = 4 - 4,8901 = -0,8901$$

3. La media de los pesos de 500 estudiantes de un colegio es 67 kg y la desviación típica 3,2 kg. Suponiendo que los pesos se distribuyen normalmente, hallar cuántos estudiantes pesan:

   a) Entre 60 kg y 65 kg.

   Se estandariza la distribución normal para calcular la probabilidad. En este caso, se tiene:

$$z_1 = \frac{60-67}{3,2} = -2,19$$

$$z_2 = \frac{65-67}{3,2} = -0,63$$

   Con la tabla x se calcula las probabilidades de cada una:

$$P_1 = 1 - 0,26435 = 0,73565$$

$$P_2 = 1 - 0,01426 = 0,98574$$

   La probabilidad es entonces:

$$P_{total} = 0,98574 - 0,73565 = 0,25$$

   Los estudiantes que pesan entre 60 kg y 65 kg son:

$$0,25 \cdot 500 = 125 \text{ personas}$$

   b) Más de 90 kg.

   Se estandariza la distribución normal para calcular la probabilidad. En este caso, se tiene:

$$z = \frac{90-67}{3,2} = 7,1875$$

   La probabilidad de obtener una observación con una puntuación z de 7 es extremadamente baja, ya que se encuentra en la cola de la distribución. De hecho, la probabilidad de obtener una puntuación z de 7 o superior en una distribución normal estándar es prácticamente cero (menor que 0.0001). Por lo tanto indica que la observación es extremadamente inusual y es probable que sea un valor atípico o una observación incorrecta.
   En conclusión se analiza que hay 0 personas con un peso de 90 kg.

   c) Menos de 64 kg.

   Se estandariza la distribución normal para calcular la probabilidad. En este caso, se tiene:

$$z = \frac{64-67}{3,2} = -0,94$$

   Con la tabla x se calcula la probabilidad:

$$P = 1 - 0,17361 = 0,82639$$

   Los estudiantes que pesan menos de 64 kg son:

$$0,82639 * 500 = 413 \text{ personas}$$

   d) 64 kg.

   Dado que la distribución normal es continua, la probabilidad de que un estudiante tenga exactamente un peso de 64 kg es cero. En otras palabras, la probabilidad de que un estudiante tenga un peso en un rango muy pequeño alrededor de 64 kg es prácticamente cero.

   e) 64 kg o menos.

   Se estandariza la distribución normal para calcular la probabilidad. En este caso, se tiene:

$$z = \frac{64-67}{3,2} = -0,94$$

Con la tabla x se calcula la probabilidad:

$$P = 1 - 0,17361 = 0,82639$$

Los estudiantes que pesan menos de 64 kg son:

$$0,82639 * 500 = 413 \text{ personas}$$

3. Una celda de carga se calibra en un ambiente a temperatura de 21°C y tiene las siguientes características de deflexión/carga. Cuando se calibra en un ambiente a temperatura de 35°C en la tabla x se pueden apreciar los cambios en las características.

Carga (Kg)	0	50	100	150	200
Deflexión(mm)	0	1	2	3	4

Cuadro I
Calibración de celda de carga a 21°C

Carga (Kg)	0	50	100	150	200
Deflexión(mm)	1.2	1.7	3.7	4.3	5.4

Cuadro II
Calibración de celda de carga a 35°C

1. Calcule la expresión matemática que mejor aproxima la salida del sensor usando el método de mínimos cuadrados matriciales. Determine la sensibilidad y la desviación del cero a 21°C y a 35°C.

   Este método se basa en encontrar la matriz de coeficientes que minimiza la distancia entre los valores medidos y los valores predichos por la ecuación. Primero, se representan los datos de calibración a 21°C y a 35°C como dos vectores columna:

$$B_{25} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, B_{35} = \begin{pmatrix} 1,2 \\ 1,7 \\ 3,7 \\ 4,3 \\ 5,4 \end{pmatrix}$$

   Luego, se construye una matriz **A** que incluye una columna de unos (para el término independiente) y una columna con las cargas medidas:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 50 \\ 1 & 100 \\ 1 & 150 \\ 1 & 200 \end{pmatrix}$$

Se resuelve el sistema de ecuaciones normales:

$$A^T A \widetilde{X} = A^T B$$

donde $\mathbf{A}^T$ es la matriz transpuesta de **A**. Dado la ecuación se calcula la transpuesta de A.

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 50 & 100 & 150 & 200 \end{pmatrix}$$

Calculamos $A^T A$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 50 & 100 & 150 & 200 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 50 \\ 1 & 100 \\ 1 & 150 \\ 1 & 200 \end{pmatrix} =$$

$$\begin{pmatrix} 5 & 500 \\ 500 & 75000 \end{pmatrix}$$

Calculamos para la celda de carga a 21 centígrados $A^T B_{21}$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 50 & 100 & 150 & 200 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 10 \\ 1500 \end{pmatrix}$$

Ahora se construye la siguiente matriz para la resolución de ecuaciones.

$$\begin{bmatrix} 5 & 500 \\ 500 & 75000 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 10 \\ 1500 \end{bmatrix}$$

Resolviendo por Gauss Jordan se tiene que:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0,02 \end{bmatrix}$$

En donde b = 0 y m= 0,02
Por lo tanto, la ecuación que mejor aproxima la salida del sensor a 21°C es:

$$y = 0,02x + 0$$

Calculamos ahora para la celda de carga a 35 centígrados $A^T B_{35}$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 50 & 100 & 150 & 200 \end{pmatrix} \cdot \begin{pmatrix} 1,2 \\ 1,7 \\ 3,7 \\ 4,3 \\ 5,4 \end{pmatrix} = \begin{pmatrix} 16,3 \\ 2180 \end{pmatrix}$$

Ahora se construye la siguiente matriz para la resolución de ecuaciones.

$$\begin{bmatrix} 5 & 500 \\ 500 & 75000 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 16,3 \\ 2180 \end{bmatrix}$$

Resolviendo por Gauss Jordan se tiene que:

$$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1,06\\0,022\end{bmatrix}$$

En donde b = 1,06 y m= 0,022
Por lo tanto, la ecuación que mejor aproxima la salida del sensor a 35°C es:

$$y = 0,022x + 1,06$$

2. Calcular el cambio de la sensibilidad a 35°C.
El cambio de sensibilidad se calcula como la diferencia de las sensibilidades de las dos temperaturas, donde se tiene:

$$\triangle S = 0,022 - 0,02 = 0,002$$

Por lo tanto, el cambio de sensibilidad a 35°C es de 0.002. Es decir, la sensibilidad del sensor aumentó en 0.002 unidades por cada unidad de carga medida, cuando la temperatura cambió de 21°C a 35°C.

3. Por lo tanto determinar los coeficientes de la desviación del cero y la sensibilidad (en unidades de mm/°C y (mm por kg)/(°C)).

La sensibilidad del sensor también puede expresarse en unidades de mm/(°C kg), dividiendo la sensibilidad en unidades de mm/kg por la diferencia de temperatura en grados Celsius.
Por lo tanto, la sensibilidad a 35°C en unidades de mm/(°C kg) es:

$$m = \frac{0,002}{35-21} = 0,001464\frac{mm/kg}{^oC}$$

La desviación del cero también puede expresarse en unidades de mm/°C, dividiendo la desviación del cero en unidades de mm por la temperatura en grados Celsius.

$$b = \frac{1,06}{35-21} = 0,07571\frac{mm}{^oC}$$

4. Con los datos de la tabla siguiente, calcule la mejor regresión lineal que la aproxima, según el método de mínimos cuadrados matriciales. Calcular el coeficiente de correlación r entre las 2 variables.

Sugerencia: Busque como se calcula la correlación, utilizando este método matricial.

49.6	47.5	46.8	50.5	45.3	48.8	48.7
104.5	101.5	99.9	106.7	97.5	101.3	102.4

Cuadro III
Conjunto de datos X,Y

Primero, se representan los datos como un vector columna:

$$B = \begin{pmatrix}104,5\\101,5\\99,9\\106,7\\97,5\\101,3\\102,4\end{pmatrix}$$

Luego, se construye una matriz **A** que incluye una columna de unos (para el término independiente) y una columna con los datos:

$$A = \begin{pmatrix}1 & 49,6\\1 & 47,5\\1 & 46,8\\1 & 50,5\\1 & 45,3\\1 & 48,8\\1 & 48,7\end{pmatrix}$$

Se resuelve el sistema de ecuaciones normales:

$$A^T A\widetilde{X} = A^T B$$

donde $\mathbf{A}^T$ es la matriz transpuesta de **A**.
Dado la ecuación se calcula la transpuesta de A.

$$A^T = \begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1\\49,6 & 47,5 & 46,8 & 50,5 & 45,3 & 48,8 & 48,7\end{pmatrix}$$

Calculamos $A^T A$

$$A^T A = \begin{pmatrix}7 & 337,3\\337,3 & 16271,49\end{pmatrix}$$

Calculamos $A^T B$

$$\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1\\49,6 & 47,5 & 46,8 & 50,5 & 45,3 & 48,8 & 48,7\end{pmatrix}.\begin{pmatrix}104,5\\101,5\\99,9\\106,7\\97,5\\101,3\\102,4\end{pmatrix} = \begin{pmatrix}713,8\\34425,18\end{pmatrix}$$

Ahora se construye la siguiente matriz para la resolución de ecuaciones.

$$\begin{bmatrix} 7 & 337,3 \\ 337,3 & 16271,49 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 713,8 \\ 34425,18 \end{bmatrix}$$

Resolviendo por Gauss Jordan se tiene que:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 23,047 \\ 1,637 \end{bmatrix}$$

En donde b = 23,047 y m= 1,637

Por lo tanto, la ecuación que mejor aproxima los datos es:

$$y = 1,637x + 23,047$$

Para el coeficiente de correlación (r) se puede calcular a partir de la matriz de covarianza y las desviaciones estándar de las variables X e Y. La fórmula para el coeficiente de correlación es:

$$r = \frac{cov(x,y)}{\sigma(x) \cdot \sigma(y)}$$

La covarianza se calcula de la siguiente manera:

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{n}$$

De esta modo se procede a calcular la media de x y y, sumando todos los valores y dividiendo por la cantidad de números.

$$\mu_x = \frac{337,2}{7} = 48,171$$

$$\mu_y = \frac{713,8}{7} = 101,9714286$$

Con los valores obtenidos se procede a calcular la covarianza y la desviación estándar:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2 = 3,119$$

$$\sigma_x = 1,766$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_y)^2 = 8,982$$

$$\sigma_y = 2,997$$

$$cov(x,y) = 5,0707$$

Por lo tanto, el coeficiente de correlación es:

$$r = \frac{5,0707}{1,766 \cdot 2,997} = 0,9579$$

5. El certificado de calibración para una resistencia estándar de laboratorio es 10.000742 Ω con 130 mV (95%). El certificado marca una incertidumbre típica sistemática de 10 mV. Estimar la incertidumbre estándar aleatoria en la resistencia.

Para calcular la incertidumbre estándar aleatoria, te tiene la siguiente formula:

$$E = \sqrt{E_a^2 + E_s^2}$$

Donde E_a es la incertidumbre aleatoria y E_s es la incertidumbre sistemática. La incertidumbre sistemática es de 10 mV, por lo que:

$$E_s = 10mV$$

Para calcular la incertidumbre aleatoria, necesitamos conocer la distribución de probabilidad de los 130 mV. Si asumimos que sigue una distribución normal, podemos utilizar el nivel de confianza del 95% para calcular la incertidumbre aleatoria.

En la distribución normal, el nivel de confianza del 95% corresponde a un factor de escala de 1.96. Por lo tanto, la incertidumbre aleatoria es:

$$E_a = \frac{130mV}{1{,}96} = 66{,}32mV$$

Sustituyendo en la fórmula de incertidumbre estándar, obtenemos:

$$E = \sqrt{66,32mV^2 + 10mV^2} = 67{,}07mV$$

Finalmente, podemos convertir la incertidumbre en Ohms utilizando la ley de Ohm:

$$R = 10{,}000742\Omega \frac{67{,}07mV}{130mV} = 5{,}15\Omega$$

Por lo tanto, la resistencia medida es:

$$R = 10{,}000742\Omega \pm 5{,}16\Omega$$

Referencias

[1] Tarea 4: Probabilidad y estadÌstica. Universidad Nacional de Colombia, 2023 [Online]. Available: https://classroom.google.com/u/0/w/NTQyNDI0ODYyMTUz/t/all?hl=es [Accessed: 29-mar-2023]