

Laboratorio de Probabilidad y Estadística

Aldo Hernández

Febrero 2025

1 Tipos de datos y Medidas de tendencia central

1. Debido a que el nombre y el área de trabajo no tienen una representación numérica, se considera que son variables cualitativas; por otra parte, la edad sí puede ser representada con números y por ello es una variable cuantitativa.

2. Calculando la media, mediana y moda de la edad:

- Usamos la fórmula para la media aritmética

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{25 + 30 + 40 + 35 + 28 + 50 + 45 + 38 + 33 + 27}{10} = \frac{351}{10} = 35.1$$

- Debido a que la cantidad de datos es par, la mediana se consigue al calcular la media de los dos valores centrales

$$\tilde{x} = \frac{33 + 35}{2} = \frac{68}{2} = 34$$

- Como no hay una edad única que se repita, el conjunto de datos es amodal.

$$\hat{x} = 25, 30, 40, 35, 28, 50, 45, 38, 33, 27$$

3. Con estas medidas de tendencia central, podemos saber que no hay empleados con la misma edad (moda), la mitad de los empleados son menores a 34 años (mediana) y la edad promedio de los empleados es de 35 años (media).

2 Medidas de dispersión

1. Para obtener la varianza, primero hay que conocer la media, esto es

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{70 + 85 + 90 + 95 + 88 + 92 + 75 + 80}{8} = \frac{675}{8} = 84.375$$

Entonces, calculamos la varianza

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{529.83}{7} = 75.69$$

Posteriormente encontramos la desviación estándar

$$s = \sqrt{s^2} = \sqrt{75.69} = 8.7$$

2. Gracias a esta información, podemos concluir que los datos están a una distancia promedio de 8.7 unidades de su media. Además, el 68.2% de los datos están en el rango de [75.675, 93.075].

3 Probabilidad y Teorema de Bayes

Para conocer la probabilidad de que al elegir un empleado al azar sea programador sabiendo que tiene conocimientos de Inteligencia Artificial, usamos el Teorema de Bayes que nos dice la probabilidad de un evento sabiendo la ocurrencia de otro. Empezamos definiendo nuestros eventos A y B junto con algunas probabilidades de interés

- $A \rightarrow$ Al elegir un empleado al azar, que sea programador.
- $B \rightarrow$ Al elegir un empleado al azar, que sea diseñador.
- $C \rightarrow$ Al elegir un empleado al azar, que tenga conocimientos de Inteligencia Artificial.
- $P(A) = 60\%$
- $P(B) = 40\%$
- $P(C/A) = 70\%$
- $P(C/B) = 30\%$

Para calcular $P(C)$ tenemos que usar el Teorema de Probabilidad Total ya que todos los empleados o son programadores o son diseñadores, y además cada empleado o sabe de Inteligencia Artificial o no tiene conocimientos sobre ello, por lo que nuestro conjunto para el evento C está compuesto por dos grupos de empleados: programadores que saben de IA y diseñadores que saben de IA

$$P(C) = P(A)P(A \cap C) + P(B)P(B \cap C) = 42\% + 12\% = 54\%$$

Entonces podemos usar el Teorema de Bayes para conocer la probabilidad de que el empleado sea programador dado que tiene conocimientos de Inteligencia Artificial

$$P(A/C) = \frac{P(C/A)P(A)}{P(C)} = \frac{70\% \cdot 60\%}{54\%} = 77.78\%$$

4 Distribuciones de Probabilidad

Según la información del problema, podemos deducir la función de distribución de probabilidad para la cantidad de defectos por lote

$$f(x) = \frac{3^x e^{-3}}{x!}; x = 0, 1, 2, \dots$$

1. Para calcular la probabilidad de que un lote tenga exactamente dos defectos, evaluamos $x = 2$ en la función

$$f(2) = \frac{3^2 e^{-3}}{2!} = \frac{9e^{-3}}{2} = 0.224 = 22.4\%$$

2. Para obtener la probabilidad de que un lote tenga al menos un defecto, usaremos el complemento de nuestro evento de interés, esto es que encontraremos la probabilidad de que tenga exactamente ningún defecto y se la restaremos a la unidad

$$\begin{aligned} P &= 1 - \bar{P} \\ &= 1 - f(0) \\ &= 1 - \frac{3^0 e^{-3}}{0!} \\ &= 1 - e^{-3} \end{aligned}$$

$$\therefore P = 0.9502 = 95.02\%$$

5 Funciones de densidad y distribución acumulativa

Deducimos la función de distribución normal

$$f(x) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(x-50)^2}{200}}; -\infty < x < \infty$$

Posteriormente normalizamos la variable aleatoria

$$z = \frac{x-50}{10}$$

1. Para calcular la probabilidad de que x sea menor que 45 usaremos la siguiente integral

$$\begin{aligned}\int_{-\infty}^{45} f(x) dx &= \frac{1}{10\sqrt{2\pi}} \int_{-\infty}^{45} e^{-\frac{(x-50)^2}{200}} dx \\ &= 0.309 \\ &= 30.9\%\end{aligned}$$

2. Para calcular la probabilidad de que x esté entre 40 y 60 usaremos la siguiente integral

$$\begin{aligned}\int_{40}^{60} f(x) dx &= \frac{1}{10\sqrt{2\pi}} \int_{40}^{60} e^{-\frac{(x-50)^2}{200}} dx \\ &= 0.683 \\ &= 68.3\%\end{aligned}$$

3. Para comprobar los resultados, usaremos la variable normalizada z

$$\begin{aligned}P(x < 45) &= P\left(\frac{x-50}{10} < \frac{45-50}{10}\right) \\ &= P\left(z < \frac{45-50}{10}\right) \\ &= 0.309 \\ &= 30.9\%\end{aligned}$$

$$\begin{aligned}P(40 < x < 60) &= P\left(\frac{40-50}{10} < \frac{x-50}{10} < \frac{60-50}{10}\right) \\ &= P(-1 < z < 1) \\ &= 1 - 2P(z > 1) \\ &= 1 - 2(0.159) \\ &= 0.682 \\ &= 68.2\%\end{aligned}$$

6 Probabilidad condicional

1. Empecemos definiendo los siguientes eventos junto con sus probabilidades

- $A \rightarrow$ Al tirar el dado en la primera tirada, el resultado sea par.
- $B \rightarrow$ Al tirar el dado en la primera tirada, el resultado sea impar.
- $C \rightarrow$ Al tirar el dado en la segunda tirada, el resultado sea par.
- $D \rightarrow$ Al tirar el dado en la segunda tirada, el resultado sea impar.
- $P(A) = P(B) = P(C) = P(D) = \frac{1}{2}$

Observamos que los eventos de interés (B y C) son independientes, es decir, la ocurrencia de uno no afecta el suceso del otro, entonces

$$P(C/B) = \frac{P(C \cap B)}{P(B)} = \frac{P(C)P(B)}{P(B)} = P(C) = \frac{1}{2} = 50\%$$

2. La primera tirada de un dado no influye en el resultado de una segunda tirada.

7 Distribución Binomial

Primero deducimos la función de probabilidad

$$f(x) = \binom{5}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{5-x}; x = 0, 1, 2, \dots, 5$$

1. Evaluamos f en 3

$$\begin{aligned} f(3) &= \binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 \\ &= (10)(0.015625)(0.5625) \\ &= 0.08789 \\ &= 8.79\% \end{aligned}$$

2. Para calcular la probabilidad, usamos el evento complementario que es el caso en el que no acierte ninguna respuesta

$$\begin{aligned} P &= 1 - f(0) \\ &= 1 - \binom{5}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^5 \\ &= (1)(1)(0.2373) \\ &= 0.2373 \\ &= 23.73\% \end{aligned}$$

8 Regla de Laplace

1. La probabilidad de obtener una bola roja es de $\frac{5}{12} = 41.67\%$ ya que hay 5 bolas rojas y un total de 12 bolas.
2. Por lo mismo, la probabilidad de sacar una bola azul es de $\frac{7}{12}$ y luego volver a sacar una bola azul sin regresar la anterior es de $\frac{6}{11}$, entonces la probabilidad P de este evento es

$$\begin{aligned}P &= \left(\frac{7}{12}\right) \left(\frac{6}{11}\right) \\&= (0.5833)(0.5454) \\&= 0.31816 \\&= 31.82\%\end{aligned}$$

9 Esperanza Matemática

1. Calculamos la esperanza matemática de la siguiente manera

$$\begin{aligned}E(x) &= (1000 - 10)(0.01) - 10(0.99) \\&= 9.9 - 9.9 \\&= 0\end{aligned}$$

2. En promedio, la persona no ganará ni perderá dinero a largo plazo.

10 Ley de los grandes números

Observamos que el número de caras es una variable aleatoria que pertenece a una distribución binomial

$$\begin{aligned}f(x) &= \binom{1000}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{1000-x} \\&= \binom{1000}{x} \left(\frac{1}{2}\right)^{1000}\end{aligned}$$

1. Esta distribución tiene un valor esperado $E(x) = n\theta = 1000(0.5) = 500$.
2. Debido a que la cantidad de repeticiones del evento es muy grande, las frecuencias relativas se acerca más al valor esperado (o teórico).