PROBABILIDAD Y ESTADISTICAS

1. Tipos de datos y Medidas de tendencia central

En una empresa se han recolectado los siguientes datos de 10 empleados:

Nombre	Edad (años)	Área de Trabajo
Ana	25	Ventas
Luis	30	Administración
Marta	40	Producción
Carlos	35	Ventas
Elena	28	Recursos Humanos
Juan	50	Producción
Sofía	45	Administración
Pedro	38	Ventas
Daniel	33	Producción
Laura	27	Recursos Humanos

1. Clasifica las variables en cualitativas o cuantitativas.

Las variables **cualitativas** son aquellas que representan una cualidad, característica o categoría, no son numéricas y no pueden medirse matemáticamente.

Las variables **cuantitativas** son aquellas que representan una cantidad y pueden medirse numéricamente.

Cualitativas

Nombre	Área de Trabajo
Ana	Ventas
Luis	Administración
Marta	Producción
Carlos	Ventas
Elena	Recursos Humanos
Juan	Producción
Sofía	Administración
Pedro	Ventas
Daniel	Producción
Laura	Recursos Humanos

Cuantitativas

Edad (años)
25
30
40
35
28
50
45
38
33
27

2. Determine la media, mediana y moda de la variable “edad”.

Los valores de edad son: $25, 30, 40, 35, 28, 32, 45, 38, 33, 27$

Media

La media se calcula como: $\bar{X} = \frac{\sum x_i}{n}$

$$\bar{X} = \frac{25 + 30 + 40 + 35 + 28 + 32 + 45 + 38 + 33 + 27}{10} = \frac{351}{10} = 35.1$$

Mediana

Ordenamos los datos de menos a mayor:

$$25, 27, 28, 30, 32, 33, 35, 38, 40, 45$$

Como hay 10 valores y son par, la mediana es el promedio de los dos valores centrales

$$Mediana = \frac{32 + 33}{2} = 32.5$$

Moda

Es el o los valores que más se repiten. Pero en esta ocasión no tenemos ningún número que se repita, por lo que no hay moda.

3. Interprete los resultados obtenidos.

- La media nos indica que la edad promedio de los empleados trabajando en la empresa es de 35.1 años.
- La mediana nos muestra que la mitad de los empleados tienen menos de 32.5 años trabajando y la otra mitad tiene más de 32.5 años.
- La moda nos indica que no hay personas que tengan la misma cantidad de años trabajando en la empresa

2. Medidas de dispersión

Dado el siguiente conjunto de datos correspondiente a las calificaciones de 8 estudiantes en un examen:

$$X = \{70, 85, 90, 95, 88, 92, 75, 80\}$$

1. Calcula las varianzas y la desviación estándar de los datos.

Primero obtenemos la media con esta fórmula:

n=8

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\bar{X} = \frac{70 + 85 + 90 + 95 + 88 + 92 + 75 + 80}{8} = \frac{675}{8} = 84.375$$

Esto nos dice que los valores están centrados alrededor de 84.375

La varianza se mide cuanto se desvían los datos de la media, con esta fórmula:

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

$$\sigma^2 = \frac{(70 - 84.375)^2 + (85 - 84.375)^2 + (90 - 84.375)^2 + (95 - 84.375)^2 + (88 - 84.375)^2 + (92 - 84.375)^2 + (75 - 84.375)^2 + (80 - 84.375)^2}{8}$$

$$\sigma^2 = \frac{206.64 + 0.39 + 31.64 + 112.89 + 13.14 + 58.14 + 87.89 + 19.14}{8} = \frac{529.87}{8} = 66.23$$

La varianza es de: 66.23

Ahora calcularemos la desviación estándar: $\sigma = \sqrt{\sigma^2}$

$$\sigma = \sqrt{66.23} = 8.13$$

2. Interprete la dispersión de los datos.
   - La varianza es de 66.23 y la desviación estándar de 8.13 esto nos dice que los valores no están tan lejos de la media de 84.375, es una dispersión moderada, ya que los datos no varían demasiado y están en un rango algo cercano al promedio.

3. Probabilidades y Teorema de Bayes

Una empresa de tecnología ha identificado que el 60% de sus empleados son programadores y el 40% son diseñadores. Se sabe que el 70% de los programadores tienen conocimientos de inteligencia artificial (IA), mientas que solo el 30% de los diseñadores tienen estos conocimientos. Si se elige un empleado al azar y se sabe que tiene conocimientos de IA, ¿Cuál es la probabilidad de que sea programador?

- $P(Prog) = 0.6$
- $P(Diseño) = 0.4$
- $P(IA|Prog) = 0.7$
- $P(IA|Diseño) = 0.3$

Queremos calcular

$$P(Prog|AI)$$

El Teorema de Bayes nos permite calcular la probabilidad de que un empleado sea programador y sabe IA, usaremos esta fórmula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(A)}$$

A,B = evento

P(A|B) = la probabilidad de A dado B

P(B|A) = la probabilidad de B dado A

P(A), P(B) = las probabilidades independientes de A y B

Sustituyendo la formula en con nuestro problema

$$P(Prog|IA) = \frac{P(AI|Prog)P(Prog)}{P(IA)}$$

Calculamos P(IA) con la regla de la probabilidad total:

$$P(IA) = P(AI|Prog)P(Prog) + P(IA|Diseño)P(Diseño)$$

$$P(IA) = (0.7 * 0.6) + (0.3 * 0.4) = 0.42 + 0.12 = 0.54$$

Ahora usamos el valor de P(IA) en la ecuación de Bayes:

$$P(Prog|IA) = \frac{(0.7)(0.6)}{0.54} = \frac{0.42}{0.54} = 0.777$$

La probabilidad de que un empleado que sabe IA sea programador es de **77.7%**

4. Distribuciones de probabilidad

Suponga que el número de defectos en un lote de producción sigue una distribución de Poisson con media $\lambda = 3$ defecto por lote.

La fórmula de Poisson es:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

1. Calcule la probabilidad de que un lote tenga exactamente 2 defectos. $P(x = 2)$

$$f(2) = \frac{3^2 e^{-3}}{2!} = \frac{9(0.0497)}{2} = 0.2236$$

Por lo tanto, hay un 22.41% de probabilidad de que un lote tenga exactamente 2 defectos

2. Calcule la probabilidad de que un lote tenga al menos 1 defecto. $P(x \geq 1)$

.

$$P(x \geq 1) = 1 - P(x = 0)$$

Calculamos P(x=0)

$$P(0) = \frac{3^0 e^{-3}}{0!} = \frac{1(0.0497)}{1} = 0.0497$$

Entonces

$$P(x \geq 1) = 1 - 0.0497 = 0.9503$$

Por lo tanto, hay un 95.03% de probabilidad de que un lote tenga al menos 1 defecto.

5. Funciones de densidad y distribución acumulada

Sea X una variable aleatoria con distribución normal de media $\mu = 50$ y una desviación estándar $\sigma = 10$.

Formula de la distribución normal estándar Z.

$$Z = \frac{X - \mu}{\sigma}$$

1. Determine la probabilidad de que X tome un valor menor que 45.$P(X < 45)$

$$Z = \frac{45-50}{10} = \frac{-5}{10} = -0.5$$

Al buscar en la tabla de la normal estándar el valor

$$P(Z < -0.5) = 0.3085$$

Por lo tanto, hay un 30.85% de probabilidad de que X sea menor a 45.

2. Determine la probabilidad de que X este entre 40 y 60. .$P(40 \leq X \geq 60$

Para **X=40:**

$$Z = \frac{40-50}{10} = -1$$

En la tabla

$$P(Z < -1) = 0.1587$$

Para **X=60:**

$$Z = \frac{60-50}{10} = 1$$

En la tabla

$$P(Z < 1) = 0.8413$$

Ahora calculamos la probabilidad

$$P(40 \leq X \leq 60) = P(Z < 1) - P(Z < -1)$$

$$= 0.8413 - 0.1587 = 0.6826$$

Por lo tanto, hay un 68.26% de probabilidad de que X este entre 40 y 60.

6. Probabilidad condicional

Un dado justo de seis caras se lanza dos veces.

1. ¿Cuál es la probabilidad de obtener un número par en el segundo lanzamiento, dado que en el primero salió un número impar?

$$\{1, 2, 3, 4, 5, 6\}$$

Impares: $\{1, 3, 5\}$

Pares: $\{2, 4, 6\}$

Aplicaremos la fórmula de probabilidad condicional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- A = obtener un número par en el segundo lanzamiento
- B = obtener un número impar en el primer lanzamiento

$$P(A|B) = P(par\ en\ el\ segundo\ dado|impar\ en\ el\ primer\ dado)$$

Probabilidad de obtener un número impar en el primer lanzamiento, como hay 3 números impares en un dado de 6 caras:

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

Probabilidad de obtener un número par en el segundo intento. El segundo lanzamiento es independiente del primero y también hay 3 números pares:

$$P(A|B) = \frac{3}{6} = \frac{1}{2}$$

2. Interprete los resultados obtenidos.

El resultado $P(A|B) = \frac{1}{2}$ significa que saber que el primer número es impar no afecta la probabilidad del segundo lanzamiento, como los lanzamientos son independientes, la probabilidad de obtener un número par en el segundo lanzamiento permanece en 50%.

7. Distribución binomial

Un examen de opción múltiple tiene 5 preguntas, cada una con 4 posibles respuestas, de las cuales solo una es correcta. Un estudiante responde al azar.

Probabilidad de acertar una pregunta:

$$\theta = \frac{1}{4} = 0.25$$

Probabilidad de fallar una pregunta:

$$1 - \theta = 1 - 0.25 = 0.75$$

Número total de intentos:

$$n = 5$$

Usamos la fórmula de la distribución binomial

$$f(x) = \binom{n}{x} \theta^{x} (1 - \theta)^{n-x}$$

1. ¿Cuál es la probabilidad de que el estudiante acierte exactamente 3 respuestas?
   $x = 3$

$$f(3) = \binom{5}{3} 0.25^{3} (1 - 0.25)^{5-3}$$

Primero calculamos el coeficiente binomial

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\,2!} = \frac{5 * 4}{2 * 1} = 10$$

Después de tener el coeficiente binomial lo sustituimos en:

$$f(3) = (10)(0.25)^{3}(1 - 0.25)^{5-3}$$

$$= (10)(0.015625)(0.5625)$$

$$= 0.0879$$

Por lo tanto, la probabilidad de que el estudiante acierte exactamente 3 respuestas es de 8.79%

2. ¿Cuál es la probabilidad de que acierte al menos una respuesta?
   $x \geq 1$

$$P(X \geq 1) = 1 - P(X = 0)$$

Calculamos P(x=0)

$$f(0) = \binom{5}{0} 0.25^0 (1 - 0.25)^5$$

$$= 1 * 1(0.75)^5$$

$$= 0.2373$$

Ahora restamos de

$$P(X \geq 1) = 1 - 0.2373$$

$$= 0.7627$$

Por lo tanto, la probabilidad de que acierte bien al menos una respuesta es de 76.27%

8. Regla de Laplace

Una urna contiene 5 bolas rojas y 7 bolas azules. Se extrae una bola al azar.

Hay 12 bolas en la urna

Usamos la fórmula de Laplace:

$$P(A) = \frac{\#\ de\ casos\ favorables}{\#\ de\ casos\ totales}$$

1. Determine la probabilidad de que la bola extraída sea roja.

$$P(roja) = \frac{\#\ de\ bolas\ rojas}{\#\ total\ de\ bolas}$$

$$P(roja) = \frac{5}{12} = 0.4166$$

Por lo tanto, la probabilidad de que la bola roja extraída sea roja es de 41.66%

2. Si se extraen dos bolas sin reemplazo, ¿Cuál es la probabilidad de que ambas sean azules?

Aquí no podremos usar Laplace ya que los eventos son dependientes, ya que el total de bolas cambia en cada extracción, usaremos probabilidad compuesta con eventos dependientes

Esta es la formula general para calcular la probabilidad de que dos eventos dependientes ocurran:

$$P(A \cap B) = P(A) * P(B|A)$$

- Donde P(A) es la probabilidad de sacar una bola azul en la primera extracción.

- P(B|A) es la probabilidad de sacar otra bola azul en la segunda extracción, dada que la primera también fue azul.

Primera extracción (A_1):

$$P(A_1) = \frac{7}{12}$$

Segunda Extracción (A_2):

En este caso como ya sacamos una bola azul, ahora solo quedan 6 bolas azules y el total de las bolas en la urna reduce a 11

$$P(A_2|A_1) = \frac{6}{11}$$

Multiplicamos ambas probabilidades

$$P(A \cap B) = \frac{7}{12} * \frac{6}{11} = 0.3181$$

Por lo tanto, la probabilidad de extraer dos bolas azules seguidas es de un 31.81%

9. Esperanza Matemática

Suponga que una persona juega una lotería donde el premio es de 1000 dólares con una probabilidad de 0.01, y el costo del boleto es de 10 dólares.

1. Calcula la esperanza matemática de la ganancia del jugador.

$$E(X) = \sum X_i P(X_i)$$

   - X_i son los posibles valores de la ganancia del jugador.
   - $P(X_i)$ es la probabilidad de cada resultado

Hay dos escenarios posibles:

Ganar el premio de 1000 dolares.

- Probabilidad: $P(ganar) = 0.01$.
- Ganancia neta: $X_1 = 100 - 10 = 990\ dolares$ (descontamos el costo del boleto).

No ganar nada

- Probabilidad: $P(perder) = 1 - 0.01 = 0.99$.
- Perdida neta: $X_2 = -10\ \ dolares$ (pierde el costo del boleto).

Sustituimos los valores en la fórmula:

$$E(X) = (990 * 0.01) + (-10 * 0.99)$$

$$E(X) = 9.90 - 9.90$$

$$E(X) = 0$$

2. Interprete el resultado obtenido.

El resultado salió 0, si juega esta lotería muchas veces, en promedio ni ganara ni perdara dinero, el precio del boleto esta equilibrado con la probabilidad de ganar, que sería un juego neutral.

10.Ley de los grandes números

Un experimento consiste en lanzar una moneda justa 1000 veces y calcular la frecuencia relativa de obtener cara.

1. ¿Cuál es el valor esperado de la frecuencia relativa de obtener cara?

$$\frac{Número\ de\ veces\ que\ ocurre\ el\ evento}{número\ total\ del\ experimento}$$

- El evento es obtener cara
- Total, de experimentos 1000

$$X \sim Bin(n = 1000, p = 0.5)$$

La esperanza matemática de una distribución binomial se calcula:

$$E(X) = n * p$$

$$E(X) = 1000 * 0.5$$

La frecuencia relativa esperada, dividimos por el número total de lanzamiento

$$\frac{E(X)}{1000} = \frac{500}{1000} = 0.5$$

El valor esperado de la frecuencia relativa de obtener cara es del 50%

2. ¿Cómo se relaciona esto con la Ley de los Grandes Números?

La ley de los Grandes Números dice que al repetir muchas veces un experimento aleatorio, la frecuencia relativa de un evento se acerca a su probabilidad teórica.

En este caso la probabilidad de obtener cara en un lanzamiento es de 0.5 y el valor esperado de la frecuencia relativa en 1000 lanzamiento también es 0.5. Aunque en 1000 intentos no siempre obtendremos exactamente 500 caras, si aumentamos los lanzamientos la frecuencia relativa se acercara cada vez más a 0.5