Laboratorio de Repaso de Probabilidad y Estadística

Probabilidad y Estadística

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. Clasifique las variables en cualitativas o cuantitativas.

Cualitativas

Área de Trabajo: $A = \{\text{Ventas}, \text{Administración}, \text{Producción}, \text{Ventas}, \text{Recursos Humanos}\}$

Nombre: $N = \{\text{Ana}, \text{Luis}, \text{Marta}, \text{Carlos}, \text{Elena}, \text{Juan}, \text{Sofía}, \text{Pedro}, \text{Daniel}, \text{Laura}\}$

Cuantitativas

Edad: $E = \{25, 30, 40, 35, 28, 50, 45, 38, 33, 27\}$

2. Determine la media, mediana y moda de la variable

Media muestral (E):

Solo es posible calcular la media para variables de tipo cuantitativas, por lo tanto, la media muestral de la edad es:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{(25 + 30 + 40 + 35 + 28 + 50 + 45 + 38 + 33 + 27)}{10} = 35.1$$

Mediana muestral (E):

Solo es posible calcular la mediana para variables de tipo cuantitativas, por lo tanto, la mediana muestral de la edad es:

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

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

Moda muestral (E):

Cada una de las edades se repiten una sola vez, por lo tanto, se trata de un conjunto de datos multimodal con:

$$m_1 = 25$$

$$m_2 = 27$$

$$m_3 = 28$$

$$m_4 = 30$$

$$m_5 = 33$$

$$m_6 = 35$$

$$m_7 = 38$$

$$m_8 = 40$$

$$m_9 = 45$$

$$m_{10} = 50$$

Moda muestral (A):

Las áreas con mayor numero de repeticiones son las de producción y ventas, con un numero de 3, por lo tanto, se trata de un conjunto de datos multimodal con modas de:

$$m_1 = producción$$

$$m_2 = ventas$$

Moda muestral (N):

Cada una de los nombres se repiten una sola vez, por lo tanto, se trata de un conjunto de datos multimodal con:

$$m_1 = Ana$$

$$m_2 = Luis$$

$$m_3 = Marta$$

$$m_4 = Carlos$$

$$m_5 = Elena$$

$$m_6 = Juan$$

$$m_7 = Sofía$$

$$m_8 = Pedro$$

$$m_9 = Daniel$$

$$m_{10} = Laura$$

3. Interprete los resultados obtenidos

Podemos concluir los siguiente:

En promedio los empleados tienen una edad de 35 años

Los departamentos mas comunes son los de producción y ventas

Es poco probable tener tocayos en la empresa (ya que en la muestra no se presenta ninguna coincidencia)

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. Calcule la varianza y la desviación estándar de los datos.

La **varianza** se calcula por medio de:

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

Donde:

$$\bar{x} = 84.375$$

$$n = 8$$

Por lo tanto:

$$s^2 = \frac{(70 - 84.375)^2 + \cdots + (80 - 84.375)^2}{7}$$

$$s^2 = 75.69$$

La **desviación estándar** se calcula como la raíz cuadrada de la varianza:

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

$$s = \sqrt{75.69}$$

$$s = 8.7$$

2. Interprete la dispersión de los datos

La desviación de los datos es relativamente baja, ya que se ve que en promedio hay una diferencia de 8.7 puntos con respecto a la media, por lo tanto, existe una consistencia en los datos, así como una variabilidad baja.

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), mientras que solo el 30% de los diseñadores poseen 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?

Se puede abordar el problema como probabilidad condicional, donde las probabilidades son:

$$A: Ser\ Programador$$

$$B: Ser\ Diseñanor$$

Por lo tanto, las probabilidades de ser solamente programador son:

$$P(A) = 0.60$$

$$P(B) = 0.40$$

Llamaremos C a la probabilidad de saber IA

El problema nos da las siguientes probabilidades:

Probabilidades de saber IA si eres Programador

$$P(C/A) = 0.70$$

Probabilidades de saber IA si eres Diseñador:

$$P(C/B) = 0.30$$

Se quiere calcular la siguiente probabilidad:

Probabilidad de ser Programador si sabes IA:

$$P(A/C)$$

Se sabe que la probabilidad condicional está dada por la siguiente formula:

$$P(A/C) = \frac{P(A \cap C)}{P(C)}$$

Pero, no se conoce $P(C)$ ni $P(A \cap C)$ directamente.

Para calcular $P(A \cap C)$:

$$P(C/A) = \frac{P(C \cap A)}{P(A)}$$

$$P(C/A)P(A) = P(C \cap A)$$

$$P(C \cap A) = (0.70)(0.60)$$

$$P(C \cap A) = P(A \cap C) = 0.42$$

Para calcular $P(C)$:

$$P(C) = P(C/A)P(A) + P(C/B)P(B)$$

$$P(C) = (0.70)(0.60) + (0.30)(0.40)$$

$$P(C) = 0.54$$

Calculamos $P(A/C)$:

$$P(A/C) = \frac{0.42}{0.54}$$

$$P(A/C) = 0.77$$

Por lo tanto, hay un 77% de probabilidad que sea programador si sabe IA

4. Distribuciones de Probabilidad

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

Se sabe que una distribución Poisson está regida por la formula:

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

1. Calcule la probabilidad de que un lote tenga exactamente 2 defectos.

$$f(2) = \frac{\boldsymbol{3^2 e^{-3}}}{\boldsymbol{2!}}$$

$$f(2) = 0.22$$

Por lo tanto, hay una probabilidad del 22% que haya exactamente 2 defectos en un lote

2. Calcule la probabilidad de que un lote tenga al menos 1 defecto.

Para ello, haremos uso de la siguiente propiedad:

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

$$P(x \leq 0) = f(0)$$

Calculamos $f(0)$:

$$f(0) = \frac{\mathbf{3^0 e^{-3}}}{\mathbf{0}!}$$

$$f(0) = 0.049$$

Reemplazamos:

$$P(x \geq 1) = 1 - 0.049$$

$$P(x \geq 1) = 0.951$$

Por lo tanto, **hay un 95.1%** de que al menos haya 1 defecto en el lote

5. Funciones de Densidad y distribución acumulativa

Sea X una variable aleatoria con distribución normal de media μ=50 y desviación estándar σ= 10.

Como se trata de una distribución normal, podemos hacer uso de las tablas, pero para ello, tendremos que normalizar la variable aleatoria X

1. Determine la probabilidad de que X tome un valor menor que 45.

$$P(x < 45)$$

$$P(\frac{X-50}{10} < \frac{45-50}{10})$$

$$P(Z < -0.5)$$

Obtenemos el valor de las tablas:

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

Por lo tanto:

$$P(x < 45) = 0.3085$$

2. Determine la probabilidad de que X esté entre 40 y 60.

$$P(40 < x < 60)$$

$$P(\frac{40-50}{10} < \frac{X-50}{10} < \frac{60-50}{10})$$

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

Obtenemos el valor de las tablas:

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

$$P(-1 < Z < 1) = 0.6826$$

Por lo tanto:

$$P(40 < x < 60) = 0.6826$$

3. Use la función de distribución acumulativa para verificar sus respuestas.

La función de probabilidad para una distribución normal es:

$$f(x) = \frac{\mathbf{1}}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \boldsymbol{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Sustituimos con los valores de la distribución:

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

Integramos la función para llegar a una expresión que nos permita calcular la probabilidad acumulada:

$$\int \frac{\mathbf{1}}{\mathbf{10}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \boldsymbol{e}^{-\frac{(x-50)^2}{200}} dx$$

$$\frac{\mathbf{1}}{\mathbf{10}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \int \boldsymbol{e}^{-\frac{(x-50)^2}{200}} dx$$

Por medio del Solver Wolfram Alpha se llegaron a los siguientes resultados:

Punto 1:

$$\frac{\mathbf{1}}{\mathbf{10}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \int_0^{45} \boldsymbol{e}^{-\frac{(x-50)^2}{200}} dx \approx 0.3085$$

Link de Wolfram Alpha

Punto 2:

$$\frac{\mathbf{1}}{\mathbf{10}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \int_{40}^{60} \boldsymbol{e}^{-\frac{(x-50)^2}{200}} dx \approx 0.6826$$

Link de Wolfram Alpha

Como se puede apreciar, los resultados coinciden con los calculados por medio de la tabla de la distribución normal

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?

En primer lugar, se especificará el espacio muestral del evento de lanzar un dado:

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

Después, se listarán 2 eventos:

$$A = \{1,3,5\}$$

$$B = \{2,4,6\}$$

Posteriormente se calcularán las probabilidades para cada evento

$$P(A) = 0.5$$

$$P(B) = 0.5$$

La probabilidad condicional está dada por la siguiente formula

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

Es decir, la probabilidad de que salga par dado que salió impar

Como A y B son independientes, (ya que no cuentan con elementos en común en sus espacios muestrales), se tiene que:

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

Sustituyendo en $P(B/A)$:

$$P(B/A) = \frac{(0.5)(0.5)}{(0.5)} = 0.5$$

2. Interprete los resultados obtenidos.

Por lo tanto**, hay un 50% de probabilidad** que salga par al lanzar por segunda vez un dado teniendo en cuenta que salió impar la primera vez (Ya que son eventos independientes)

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.

Se trata de una distribución de probabilidad binomial, con $\boldsymbol{\theta = 0.25,}$ ya que las probabilidades de éxito por pregunta son de ¼

$$f(x) = \binom{\mathbf{5}}{\boldsymbol{x}}(\mathbf{0.25})^{x}(\mathbf{0.75})^{\mathbf{5}-x}$$

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

Para ello calculamos $f(3)$

$$f(3) = \binom{\mathbf{5}}{\mathbf{3}}(\mathbf{0.25})^{3}(\mathbf{0.75})^{2}$$

$$f(3) = 0.087$$

Por lo tanto, hay un 8.7% de acertar exactamente 3 preguntas de las 5

2. ¿Cuál es la probabilidad de que acierte al menos una respuesta?

Para ello, hacemos uso de una propiedad de la probabilidad acumulada:

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

$$P(x \leq 0) = f(0)$$

$$f(0) = \binom{\mathbf{5}}{\mathbf{0}}(\mathbf{0.25})^{\mathbf{0}}(\mathbf{0.75})^{\mathbf{5}}$$

$$f(0) = 0.237$$

$$P(x \geq 1) = 1 - 0.237$$

$$P(x \geq 1) = 0.762$$

Por lo tanto, hay un 76.2% de acertar por lo menos una pregunta

8. Regla de Laplace

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

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

La probabilidad de que la bola sea de color rojo estará dada por:

$$P(R) = \frac{5}{12} = 0.416$$

Es decir, hay un 41.6% de probabilidad de que sea roja

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

Aquí se tiene que ver como un experimento por etapas:

$$\frac{7}{12} \times \frac{6}{11}$$

En la primera etapa se tienen 7 bolas azules de las 12 que hay

En la segunda etapa se tiene 6 bolas azules de las 11 (ya que la bola anterior debió ser azul), entonces:

$$(0.583)(0.5454) = 0..3179$$

Por lo tanto, hay un 31.79% de que ambas bolas sean azules

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. Calcule la esperanza matemática de la ganancia del jugador.

$$E(x) = (P(ganar) * ganancia) + P(perder * costo)$$

$$E(x) = (0.01)(990) + (0.99)(-10)$$

$$E(x) = 9.9 - 9.9$$

$$E(x) = 0$$

2. Interprete el resultado obtenido.

La persona en promedio no ganará ni perderá dinero, ya que su esperanza matemática es 0

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?

Definimos el espacio muestral:

$$M = \{C, A\}$$

Definimos 2 eventos:

$$C = \{C\}$$

$$A = \{A\}$$

La probabilidad para cada evento es de:

$$P(C) = 1/2$$

$$P(A) = 1/2$$

Definimos la frecuencia relativa de obtener cara como el número de caras obtenidas dividido entre el número total de lanzamientos (1000). Esto se expresa como:

$$F_{RC} = \frac{\#\ de\ caras}{1000}$$

El valor esperado de la frecuencia relativa es simplemente la probabilidad de obtener cara en un solo lanzamiento, ya que, al realizar muchos lanzamientos, la frecuencia relativa debería aproximarse a la probabilidad de obtener cara en cada lanzamiento.

Por lo tanto, el valor esperado de la frecuencia relativa de obtener cara es:

$$E(F_{RC}) = P(C) = 0.5$$

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

La Ley de los Grandes Números establece que, a medida que el número de repeticiones de un experimento aleatorio aumenta, la frecuencia relativa de un evento (en este caso, obtener cara) se aproxima a la probabilidad teórica de ese evento.