Probabilidad y Estadística

El propósito de este laboratorio es familiarizarse con conceptos fundamentales de probabilidad y estadística, a través de problemas prácticos relacionados con los subtemas estudiados.

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.

   Variables cualitativas: Nombre y área de trabajo
   Variables cuantitativas: Edad

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

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

   $\bar{x} = 35.1$

   Mediana = X25, X27, X28, X30, 33, 35, X38, X40, X45, X50
   Mediana = 34 (promedio de 33 y 35)

   Moda: Como ningún dato se repite, en conjunto de datos es amodal.

3. Interprete los resultados obtenidos.

   En el conjunto de datos estudiado, se puede decir que el grupo de edades tiene un distribución relativamente equilibrada, pero con una ligera tendencia hacia los valores más altos, dado que la media (35.1) es un poco mayor a la mediana (34). Sobre la moda, se puede decir que no hay una edad predominante y que las edades están distribuidas de manera dispersa.

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 con la siguiente fórmula:

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

   $\sigma^2 = 66.23$

   La desviación estándar se calcula con la siguiente fórmula:

   $$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

   $\sigma = 8.13$

2. Interprete la dispersión de los datos.

   Gracias a la desviación estándar, podemos inferir que los datos, en promedio, están alejados 8.13 puntos del promedio de 84.37 (ya sea para arriba o para abajo).

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 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 = ser programador
D = ser diseñador
IA = tener conocimientos sobre la IA

Usando el teorema de Bayes, tenemos que:

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

donde P(IA) se calcula con la ley de la probabilidad total:

$$P(IA) = P(IA|P)P(P) + P(IA|D)P(D)$$

Sustituimos valores y resolvemos el problema:

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

P(IA) = 0.54
Regresamos al teorema de Bayes y tenemos que:

$$P(P|IA) = \frac{(0.7*0.6)}{0.54} = 0.7778$$

Por lo tanto el resultado es que hay un 77.78% probabilidades de que sea programador.

4. Distribuciones de probabilidad

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

1. Calcule la probabilidad de que un lote tenga exactamente 2 defectos.
   Aplicamos la fórmula con los datos dados:
   $$P(X=2) = \frac{3^2 e^{-3}}{2!}$$
   P(X = 2) = 0.2240
   Por lo tanto, hay un 22.40% de que salgan dos defectos.

2. Calcule la probabilidad de que un lote tenga al menos 1 defecto.
   $$P(X \geq 1) = 1 - P(X=0)$$
   $$P(X=0) = \frac{\lambda^0 e^{-\lambda}}{0!}$$
   $$P(X \geq 1) = 1 - e^{-\lambda}$$
   $$P(X \geq 1) = 1 - e^{-3}$$
   $$P(X \geq 1) = 0.9502$$
   Por lo tanto hay al menos 95.02% de que al menos uno sea defectuoso.

5. Funciones de densidad y distribución acumulativa

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

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

   Para resolver el problemas hay que hacer uso de la tabla de la distribución normal, pero como solo sirve cuando N(0,1), hay que normalizar los datos, se hace con la siguiente fórmula:
   $$Z = \frac{X - \mu}{\sigma}$$
   Sustituimos datos:
   $$Z = \frac{45-50}{10} = \frac{-5}{10} = -0.5$$
   Buscamos el valor anterior en la tabla y nos da que el valor de es:
   $$P(Z < -0.5) \approx 0.3085$$
   $$P(X < 45) = 0.3085$$
   Por lo tanto, hay un 30.85% de que tome un valor menor que 45.

2. Determine la probabilidad de que X esté entre 40 y 60.
   Seguimos el mismo procedimiento anterior:
   $$P(40 < X < 60)$$
   $$Z = \frac{X - \mu}{\sigma}$$
   $$Z = \frac{40-50}{10} = \frac{-10}{10} = -1$$
   $$Z = \frac{60-50}{10} = \frac{10}{10} = 1$$
   $$P(Z < -1) \approx 0.1587$$
   $$P(Z < 1) \approx 0.8413$$
   $$P(40 < X < 60) = P(Z < 1) - P(Z < -1)$$
   $$= 0.8413 - 0.1587 = 0.6826$$
   $$P(40 < X < 60) = 0.6826$$
   Por lo tanto, hay un 68.26% de posibilidades de que tome un valor entre 40 y 60.

3. Use la función de distribución acumulativa para verificar sus respuestas.
   $$P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$
   $$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1)$$
   $$P(40 < X < 60) = \Phi(1) - \Phi(-1)$$
   $$= 0.8413 - 0.1587$$
   $$= 0.6826$$

Utilizando la FDA de la normal estándar:

$$P(Z < -0.5) = \Phi(-0.5)$$

$$\Phi(-0.5) \approx 0.3085$$

Por lo tanto:

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

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?

   Probabilidad de que salga impar en el primer intento:

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

   Probabilidad de que salga par en el segundo intento:

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

   Y como los eventos son independientes, se puede simplificar a:

   $$P(B \mid A) = P(B) = \frac{1}{2}$$

2. Interprete los resultados obtenidos.
   Hay un 50% de probabilidades de que en el segundo intendo salga par, dado que en el segundo salió impar.

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.

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

   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

   $$P(X = 3) = \binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2$$

   Por lo tanto, la probabilidad para que acierte exactamente tres respuestas es del 8.75%.

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

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

   $$P(X = 0) = \binom{5}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^5$$

   $$P(X \geq 1) = 1 - \left(\frac{3}{4}\right)^5$$

   Por lo tanto, la probalidad para que acierte al menos una respuesta es del 76.26%

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.

   $$P(R) = \frac{5}{5+7} = \frac{5}{12}$$

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

   $$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 \mid A_1)$$

   $$P(A_1 \cap A_2) = \frac{7}{12} \times \frac{6}{11}$$

   $$= \frac{42}{132} = \frac{7}{22}$$

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.

   Si gana, tiene una probabilidad del 0.01 con 990 dolares de ganancia, si pierde, serían -10 dolares de ganancia con una probabilidad de 0.99, por lo tanto la esperanza se calcularia de la siguiente manera:

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

   E(X) = 0
   La esperanza matemática es $0, lo que significa que, en promedio, el jugador no gana ni pierde dinero.

2. Interprete el resultado obtenido.
   Una esperanza matemática de \$0 sugiere que, a largo plazo, el jugador no tiene una ganancia ni una pérdida esperada. Esto implica que el juego es justo en términos de valor esperado. Se resalta lo de a largo plazo, porque si lo compra una sola vez y no gana, perderá 10 dólares.

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?
   Esta sería la probabilidad de que salga cara:
   $$E(f) = \frac{1}{2}$$
2. ¿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 experimentos aumenta, la frecuencia relativa de un evento se aproxima a su probabilidad teórica. Por lo tanto, si realizamos el evento en múltiples ocasiones, el resultado convergerá a un medio:
   $$\lim_{n\to\infty} \frac{X}{n} = P(C) = \frac{1}{2}$$