Apuntes de Estadística y Probabilidad

Fórmulas Claves de Estadística y Probabilidad

0.1 Probabilidad de la unión de dos eventos

Uso: Calcula la probabilidad de que ocurra al menos uno de dos eventos.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

donde:

- $P(A)$ y $P(B)$ son las probabilidades individuales de los eventos A y B.
- $P(A \cap B)$ es la probabilidad de que ambos eventos ocurran simultáneamente.

Ejemplo: En una empresa, el 40% de los empleados tienen coche propio y el 35% usa transporte público. Se sabe que el 15% tiene coche y también usa transporte público.

$$P(A) = 0.40, \quad P(B) = 0.35, \quad P(A \cap B) = 0.15$$

Calculamos la probabilidad de que un empleado tenga coche o use transporte público:

$$P(A \cup B) = 0.40 + 0.35 - 0.15 = 0.60$$

Por lo tanto, la probabilidad es 0.60.

0.2 Probabilidad condicional

Uso: Calcula la probabilidad de que ocurra un evento dado que otro ya ha ocurrido.

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

donde:

- $P(A|B)$ es la probabilidad de que ocurra A dado que ocurrió B.
- $P(A \cap B)$ es la probabilidad conjunta de que ocurran A y B.
- $P(B)$ es la probabilidad de B.

Ejemplo: En un gimnasio, el 70% de los socios asisten regularmente y el 50% toma suplementos. Se sabe que el 40% asiste regularmente y también toma suplementos.

$$P(A) = 0.70, \quad P(B) = 0.50, \quad P(A \cap B) = 0.40$$

Queremos calcular $P(A|B)$ (probabilidad de que alguien asista regularmente dado que toma suplementos):

$$P(A|B) = \frac{0.40}{0.50} = 0.80$$

Por lo tanto, la probabilidad es 0.80.

0.3 Teorema de la probabilidad total

Uso: Calcula la probabilidad de un evento considerando varios casos posibles.

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

donde:

- $B_1, B_2, \ldots, B_n$ son eventos que forman una partición del espacio muestral.
- $P(A|B_i)$ es la probabilidad de que ocurra A dado que ocurrió B_i.
- $P(B_i)$ es la probabilidad de cada B_i.

Ejemplo: Un parque tiene dos entradas, una por el norte y otra por el sur. El 60% de los visitantes entran por el norte y el 40% por el sur. Se sabe que el 30% de los que entran por el norte llevan sombrero y el 50% de los que entran por el sur también lo llevan.

Calculamos la probabilidad de que un visitante elegido al azar lleve sombrero:

$$P(H) = P(H|N)P(N) + P(H|S)P(S)$$

Sustituyendo valores:

$$P(H) = (0.30 \times 0.60) + (0.50 \times 0.40)$$

$$P(H) = 0.18 + 0.20 = 0.38$$

Por lo tanto, la probabilidad de que un visitante lleve sombrero es 0.38.

0.4 Regla de Bayes

Uso: Calcula la probabilidad de una causa dado un resultado.

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

donde:

- $P(B|A)$ es la probabilidad de que la causa sea B dado que ocurrió A.
- $P(A|B)$ es la probabilidad de obtener A si se dio B.
- $P(B)$ es la probabilidad de que ocurra B.
- $P(A)$ es la probabilidad total de A.

Ejemplo: En una fábrica, el 65% de los productos provienen de la línea A y el 35% de la línea B. Se sabe que el 80% de los productos de la línea A pasan el control de calidad, mientras que el 90% de los productos de la línea B también lo pasan.

Queremos calcular la probabilidad de que un producto que ha pasado el control de calidad provenga de la línea B.

Datos:

$$P(A) = 0.65, \quad P(B) = 0.35, \quad P(Q|A) = 0.80, \quad P(Q|B) = 0.90$$

Calculamos la probabilidad de que un producto pase el control de calidad:

$$P(Q) = P(Q|A)P(A) + P(Q|B)P(B)$$

$$P(Q) = (0.80 \times 0.65) + (0.90 \times 0.35) = 0.52 + 0.315 = 0.835$$

Ahora aplicamos la regla de Bayes para calcular $P(B|Q)$:

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

$$P(B|Q) = \frac{(0.90 \times 0.35)}{0.835} = \frac{0.315}{0.835} \approx 0.377$$

Por lo tanto, la probabilidad de que un producto que ha pasado el control de calidad provenga de la línea B es 0.377.

0.5 Función de Densidad de Probabilidad (FDP)

Uso: Se usa para modelar la distribución de probabilidad de una variable aleatoria continua.

$$f(x) = \begin{cases} c(1-x^2), & 0 \leq x \leq 1 \\ 0, & \text{en otro caso} \end{cases}$$

donde:

- $f(x)$ es la función de densidad de probabilidad.
- c es una constante de normalización.
- x es la variable aleatoria.

Ejemplo: En un análisis de velocidad de automóviles en una autopista, se modela la distribución de velocidades con $f(x) = c(1-x^2)$. Para encontrar c, se resuelve:

$$\int_0^1 c(1-x^2)\, dx = 1$$

Resolviendo la integral:

$$c = \frac{3}{2}$$

—

0.6 Esperanza Matemática y Varianza

Uso: La esperanza ($E(X)$) representa el valor esperado de una variable aleatoria, y la varianza ($V(X)$) mide su dispersión.

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$V(X) = E(X^2) - [E(X)]^2$$

donde:

- $E(X)$ es la media esperada.
- $V(X)$ es la varianza.
- $f(x)$ es la función de densidad de probabilidad.

Ejemplo: Se modela la altura de los estudiantes en una escuela con una función de densidad de probabilidad. Resolviendo las integrales:

$$E(X) = \frac{3}{8}, \quad V(X) = \frac{19}{320}$$

—

0.7 Distribución Exponencial

Uso: Se usa para modelar tiempos de espera entre eventos en un proceso aleatorio.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

donde:

- λ es la tasa de ocurrencia de eventos.
- x es el tiempo hasta que ocurra el evento.

Ejemplo: El tiempo entre llamadas a un servicio de emergencias sigue una distribución exponencial con $\lambda = 3$ llamadas por hora. La probabilidad de que pasen más de 2 horas sin recibir una llamada es:

$$P(X > 2) = e^{-3\times 2} = e^{-6} \approx 0.002478$$

—

0.8 Distribución de Poisson

Uso: Modela el número de eventos en un intervalo fijo de tiempo o espacio.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

donde:

- λ es la tasa media de ocurrencias en un intervalo.

- k es el número de eventos en el intervalo.

Ejemplo: Un hospital recibe en promedio 10 pacientes por hora en urgencias. La probabilidad de que lleguen exactamente 7 pacientes en una hora es:

$$P(X=7) = \frac{10^7 e^{-10}}{7!} \approx 0.0902$$

—

0.9 Distribuciones Conjuntas e Independencia

Uso: Se usa para analizar la relación entre dos variables aleatorias.

$$f(x,y) = \begin{cases} \frac{2-x-y}{8}, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{en otro caso} \end{cases}$$

Si $f(x,y) \neq f_X(x) f_Y(y)$, entonces las variables no son independientes.

Ejemplo: En un estudio sobre la correlación entre la altura y el peso de personas, se obtiene la función de densidad conjunta. Calculando $f_X(x)$ y $f_Y(y)$, se concluye que no son independientes.

—

0.10 Intervalo de Confianza para la Media

Uso: Se usa para estimar un rango en el que se espera que se encuentre la media poblacional.

$$IC = \bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

donde:

- $\bar{X}$ es la media muestral.
- σ es la desviación estándar poblacional.
- n es el tamaño de la muestra.
- $Z_{\alpha/2}$ es el valor crítico de la distribución normal.

Ejemplo: Se quiere estimar el tiempo medio de reparación de un automóvil con una desviación estándar de 20 minutos. Con una muestra de 50 reparaciones, el intervalo de confianza al 95% es:

$$IC = \bar{X} \pm 1.96 \frac{20}{\sqrt{50}}$$

—

0.11 Aproximación Normal a la Distribución Muestral

Uso: Para muestras grandes, la distribución de la media muestral se aproxima a una normal.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

donde:

- $\bar{X}$ es la media muestral.
- μ es la media poblacional.
- σ es la desviación estándar poblacional.
- n es el tamaño de la muestra.

Ejemplo: Se toma una muestra de $n = 25$ de una distribución normal con media $\mu = 50$ y desviación estándar $\sigma = 5$. Queremos calcular la probabilidad de que la media muestral sea mayor a 52.

$$P(X > 52) = P\left(Z > \frac{52 - 50}{5/\sqrt{25}}\right)$$

$$P(Z > 2) \approx 0.0228$$

0.12 Intervalo de Confianza para la Varianza

Uso: Se utiliza para estimar un rango en el que es probable que se encuentre la varianza poblacional σ^2 basándose en una muestra.

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right)$$

donde:

- n es el tamaño de la muestra.
- S^2 es la varianza muestral.
- $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son los valores críticos de la distribución χ^2.

Ejemplo: Se mide la cantidad de radiación en el agua en 10 muestras, obteniendo una varianza muestral de 34. Con un nivel de confianza del 90%, los valores críticos de χ^2 para 9 grados de libertad son 16.92 y 3.325.

$$\left(\frac{9 \times 34}{16.92}, \frac{9 \times 34}{3.325}\right) = (18.10, 92.08)$$

Por lo tanto, el intervalo de confianza es $(18.10, 92.08)$.

—

0.13 Intervalo de Confianza para la Diferencia de Medias

Uso: Se emplea para estimar la diferencia entre dos medias poblacionales.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2} \cdot SE$$

donde:

- $\bar{x}_1, \bar{x}_2$ son las medias muestrales de los dos grupos.
- $t_{1-\alpha/2}$ es el valor crítico de la distribución t.
- SE es el error estándar, dado por:

$$SE = \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

- S_p^2 es la varianza combinada:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Ejemplo: Se comparan los ingresos de dos grupos con los siguientes datos:

$$\bar{x}_1 = 3565, S_1 = 150, n_1 = 25, \quad \bar{x}_2 = 3280, S_2 = 170, n_2 = 12$$

Tras calcular S_p^2, obtenemos el intervalo:

$$(173.51, 396.49)$$

Esto indica que con un 95% de confianza, la diferencia de medias está en ese rango.

—

0.14 Prueba F para Comparación de Varianzas

Uso: Se usa para comparar las varianzas de dos poblaciones mediante la distribución F de Fisher.

$$F = \frac{S^2_{\text{mayor}}}{S^2_{\text{menor}}}$$

donde:

- S_A^2, S_B^2 son las varianzas muestrales.
- F sigue una distribución con $n_A - 1$ y $n_B - 1$ grados de libertad.

Ejemplo: Se comparan dos tipos de renta fija con:

$$S_A^2 = 125.25, \quad S_B^2 = 638.5, \quad n_A = n_B = 17$$

Calculamos el estadístico:

$$F = \frac{638.5}{125.25} = 5.1$$

Comparando con los valores críticos, se concluye que las varianzas son significativamente diferentes.

—

0.15 Distribución Normal de la Media Muestral

Uso: Se usa para calcular probabilidades sobre la media de una muestra de tamaño n.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

donde:

- $\bar{X}$ es la media muestral.
- μ es la media poblacional.
- σ es la desviación estándar poblacional.
- n es el tamaño de la muestra.

Ejemplo: Se empaquetan bolsas de azúcar en lotes de 100 unidades. Se sabe que $\mu = 500$g, $\sigma = 35$g. Queremos calcular la probabilidad de que la media de un lote sea menor de 495g.

Calculamos el error estándar:

$$SE = \frac{35}{\sqrt{100}} = 3.5$$

El estadístico Z es:

$$Z = \frac{495 - 500}{3.5} = -1.43$$

De la tabla normal:

$$P(Z < -1.43) = 0.0764$$

Por lo tanto, la probabilidad es 7.64%.

—

0.16 Intervalo de Confianza para la Media

Uso: Se emplea para estimar un rango en el que se espera que se encuentre la media poblacional.

$$\left(\mu - Z_{1-\alpha/2} \cdot SE, \mu + Z_{1-\alpha/2} \cdot SE\right)$$

donde:

- $Z_{1-\alpha/2}$ es el valor crítico de la normal estándar.
- SE es el error estándar.

Ejemplo: Se desea estimar la media del peso de las bolsas de azúcar con un nivel de confianza del 95%.

$$Z_{0.975} = 1.96, \quad SE = 3.5$$

Calculamos los límites:

$$500 - 1.96 \times 3.5 = 493.14$$

$$500 + 1.96 \times 3.5 = 506.86$$

El intervalo de confianza es $(493.14, 506.86)$.
—

0.17 Probabilidad de un Peso Total Mayor a un Valor Dado

Uso: Se calcula la probabilidad de que la suma de una muestra de valores supere un umbral.

$$Z = \frac{X - n\mu}{SE_{\text{total}}}$$

donde:

- X es el valor total a evaluar.
- $SE_{\text{total}} = n \times SE$.

Ejemplo: Se quiere saber la probabilidad de que una caja de 100 bolsas pese más de 51 kg.

$$Z = \frac{51000 - 50000}{350} = 2.86$$

Buscando en la tabla normal:

$$P(Z > 2.86) = 0.0021$$

Esto indica que la probabilidad es **0.21%**.