Universidad

Nacional Autónoma de México

Facultad de Ciencias

Estadística

Integrantes:
Yonathan Berith Jaramillo Ramírez. 419004640

Profesor: Ruth Selene Fuentes García
Ayudante: Carlos Alberto Arriaga Solórzano

10 Octubre, 2023

Estadística descriptiva

Medidas muestrales de tendencia central

Las medidas de tendencia central son estadísticas que tratan de resumir en un solo valor un conjunto de datos numéricos. Estas medidas son útiles para proporcionar una visión general de la distribución de los datos, permitiendo hacer comparaciones y análisis más simples que trabajar con todo el conjunto de datos original. Las medidas de tendencia central más comunes son la media aritmética, la mediana y la moda.

La media muestral

La **media muestral** ($\bar{x}$) es una estadística que se calcula como la suma de todas las observaciones en una muestra, dividida por el número total de observaciones en la muestra. Matemáticamente, la media muestral se define como:

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

donde n es el número de observaciones en la muestra y x_i es el valor de la i-ésima observación, $\bar{x}$ tambien es conocído como μ

La mediana muestral

La **mediana muestral** es una medida de tendencia central que divide una muestra de datos en dos partes iguales. Para calcular la mediana muestral:

1. Ordene los elementos de la muestra de manera ascendente, es decir, $x_1 \leq x_2 \leq \ldots \leq x_n$.
2. Si n es impar, la mediana es el valor que se encuentra en la posición central:
$$\text{Mediana} = x_{\frac{n+1}{2}} \tag{2}$$
3. Si n es par, la mediana es el promedio de los dos valores centrales:
$$\text{Mediana} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \tag{3}$$

Porcentil

Un **porcentil** es una medida estadística que describe la posición relativa de un valor dentro de una muestra o población ordenada. Específicamente, el p-ésimo porcentil de un conjunto de datos es el valor por debajo del cual cae un porcentaje p de las observaciones en el conjunto de datos.

Para calcular el p-ésimo porcentil de una muestra:

1. Ordene los elementos de la muestra en orden ascendente, es decir, $x_1 \leq x_2 \leq \ldots \leq x_n$.
2. Encuentre el índice k mediante la siguiente fórmula:

$$k = \frac{p(n-1)+1}{100} \tag{4}$$

3. Si k es un número entero, el p-ésimo porcentil es x_k.
4. Si k no es un número entero, redondee k al entero más cercano para encontrar el p-ésimo porcentil más cercano en los datos.

Medidas de dispersión

Las medidas de dispersión son estadísticas que describen la variabilidad o el "esparcimiento" de un conjunto de datos. A diferencia de las medidas de tendencia central, que ofrecen un valor único que representa de alguna manera un "centro" para el conjunto de datos, las medidas de dispersión brindan una idea de cuán "esparcidos" están los datos alrededor de ese centro. Algunas de las medidas de dispersión más comunes son la varianza, la desviación estándar, el rango y el rango intercuartílico.

Varianza

La varianza es un indicador cuantitativo de la dispersión de los datos alrededor de la media. Se calcula como el promedio de las diferencias al cuadrado entre cada observación y la media aritmética del conjunto. Matemáticamente, la varianza de una muestra s^2 se define como:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{5}$$

Desviación Estándar

La desviación estándar es simplemente la raíz cuadrada de la varianza. Proporciona una medida de dispersión que está en las mismas unidades que los datos originales.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{6}$$

Rango Muestral

El rango es la diferencia entre el valor máximo y el valor mínimo del conjunto de datos. Es la medida más simple de dispersión pero es sensible a los valores extremos o atípicos.

$$\text{Rango} = x_{\text{max}} - x_{\text{min}} \tag{7}$$

Rango Intercuartílico (IQR)

El rango intercuartílico es una medida de dispersión más robusta que el rango, ya que no es sensible a valores extremos. Se calcula como la diferencia entre el tercer cuartil ($Q3$) y el primer cuartil ($Q1$).

$$\text{IQR} = Q3 - Q1 \tag{8}$$

Estadísticas y distribuciones muestrales

Estimación puntual

La estimación puntual es un concepto en estadística inferencial que implica utilizar los datos de una muestra para calcular un único valor (o "punto") que se considera como una buena aproximación del parámetro poblacional desconocido. En otras palabras, un estimador puntual es una estadística (una función de los datos observados) que proporciona un valor único para un parámetro desconocido de la población.

Momento muestral

Un "momento muestral" es un concepto estadístico que se utiliza para describir diversas características de una muestra de datos. Los momentos son útiles para entender la forma de la distribución subyacente a partir de la cual se extrae la muestra. Los momentos muestrales son calculados usando los datos de la muestra, a diferencia de los "momentos poblacionales", que se calculan usando todos los datos en una población completa.

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

Aquí, n es el número de observaciones en la muestra, x_i es el valor de la i-ésima observación, y $\bar{x}$ es la media muestral. Cuando $k = 1$, el momento se simplifica a la media de los datos; cuando $k = 2$, el momento se relaciona con la varianza muestral.

Método de Momentos en Estadística

El Método de Momentos es una técnica estadística utilizada para estimar los parámetros de una distribución de probabilidad. Se basa en igualar los momentos muestrales a los momentos teóricos de la distribución en cuestión para resolver las ecuaciones resultantes con respecto a los parámetros que se quieren estimar.

Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución con función de densifdad $f(x;\theta)$.

Momentos Muestrales y Teóricos

Matemáticamente, el k-ésimo momento muestral se denota por M_k y se calcula cómo:

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{9}$$

Matemáticamente, el k-ésimo momento poblacional se denota por μ_k y se calcula como:

$$E(X_i^k) \tag{10}$$

Y el k-ésimo momento teórico se expresa generalmente en términos de los parámetros de la distribución. Por ejemplo, el primer y segundo momento de una distribución normal son μ y $\mu^2 + \sigma^2$, respectivamente.

Pasos del Método de Momentos

Para llevar a cabo el Método de Momentos:

1. Se calculan los momentos muestrales hasta el orden que sea necesario para estimar los parámetros deseados.
2. Se igualan estos momentos muestrales a los momentos teóricos de la distribución.
3. Se resuelven las ecuaciones resultantes para obtener estimaciones de los parámetros.

Estimador de Máxima Verosimilitud para θ

DISTRIBUCIÓN BERNULLI

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de densidad de probabilidad $f(x;\theta)$, donde θ es un parámetro desconocido. La función de verosimilitud $L(\theta; x_1, x_2, \ldots, x_n)$ se define como:

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

El Estimador de Máxima Verosimilitud (MLE) de θ, denotado como $\hat{\theta}_{MLE}$, es el valor de θ que maximiza la función de verosimilitud $L(\theta)$. Matemáticamente, $\hat{\theta}_{MLE}$ se encuentra resolviendo:

$$\frac{dL(\theta)}{d\theta} = 0$$

y asegurando que la segunda derivada sea negativa para confirmar un máximo local.

Estimación por intervalos

Temas básicos a dominar

Propiedades de los logaritmos

1. **Identidad**: $\log_b(b) = 1$
2. **Logaritmo Base 1**: $\log_1(x)$ es indefinido.
3. **Logaritmo de 1**: $\log_b(1) = 0$
4. **Logaritmo de 0**: $\log_b(0)$ es indefinido para $b > 0$.
5. **Cambio de Base**: $\log_b(a) = \frac{\log_c(a)}{\log_c(b)}$

Propiedades Operacionales de los Logaritmos

1. **Producto**: $\log_b(xy) = \log_b(x) + \log_b(y)$
2. **Cociente**: $\log_b(x/y) = \log_b(x) - \log_b(y)$
3. **Potencia**: $\log_b(x^n) = n \log_b(x)$
4. **Raíz**: $\log_b(\sqrt[n]{x}) = \frac{1}{n} \log_b(x)$

Propiedades Especiales de los Logaritmos

1. **Logaritmo Recíproco**: $\log_b(1/x) = -\log_b(x)$
2. **Inversa**: $b^{\log_b(x)} = x$ y $\log_b(b^x) = x$
3. **Logaritmo Natural y Exponencial**: $\log_e(x) = \ln(x)$ y $e^{\ln(x)} = x$

Propiedades de los exponentes

1. **Producto de Bases Iguales**: $a^m \times a^n = a^{m+n}$

2. **Cociente de Bases Iguales**: $\frac{a^m}{a^n} = a^{m-n}$, $a \neq 0$
3. **Potencia de una Potencia**: $(a^m)^n = a^{m \times n}$
4. **Potencia de un Producto**: $(a \times b)^n = a^n \times b^n$
5. **Potencia de un Cociente**: $\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$, $b \neq 0$
6. **Exponente Cero**: $a^0 = 1$, $a \neq 0$
7. **Exponente Negativo**: $a^{-n} = \frac{1}{a^n}$, $a \neq 0$
8. **Exponente Fraccional**: $a^{1/n} = \sqrt[n]{a}$, $a > 0, n > 0$
9. **Exponente Uno**: $a^1 = a$

Derivadas básicas

$$\frac{d}{dx}(c) = 0$$
$$\frac{d}{dx}(x^n) = nx^{n-1}$$
$$\frac{d}{dx}(e^x) = e^x$$
$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$$
$$\frac{d}{dx}(\sin(x)) = \cos(x)$$
$$\frac{d}{dx}(\cos(x)) = -\sin(x)$$
$$\frac{d}{dx}(\tan(x)) = \sec^2(x)$$
$$\frac{d}{dx}(\cot(x)) = -\csc^2(x)$$
$$\frac{d}{dx}(\sec(x)) = \sec(x)\tan(x)$$
$$\frac{d}{dx}(\csc(x)) = -\csc(x)\cot(x)$$
$$\frac{d}{dx}(\arcsin(x)) = \frac{1}{\sqrt{1-x^2}}$$
$$\frac{d}{dx}(\arccos(x)) = -\frac{1}{\sqrt{1-x^2}}$$
$$\frac{d}{dx}(\arctan(x)) = \frac{1}{1+x^2}$$

Reglas de Derivación

Regla de la suma

$$\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}$$

Regla del producto

$$\frac{d}{dx}(u \times v) = u\frac{dv}{dx} + v\frac{du}{dx}$$

Regla del cociente

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

Regla de la cadena

$$\frac{d}{dx}(f(g(x))) = f'(g(x)) \times g'(x)$$

Integrales Indefinidas Básicas

$$\int k\,dx = kx + C$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1$$

$$\int e^x\,dx = e^x + C$$

$$\int \frac{1}{x}\,dx = \ln|x| + C$$

$$\int \sin(x)\,dx = -\cos(x) + C$$

$$\int \cos(x)\,dx = \sin(x) + C$$

$$\int \tan(x)\,dx = \ln|\sec(x)| + C$$

$$\int \sec^2(x)\,dx = \tan(x) + C$$

Reglas de Integrales

Integral de una suma

$$\int (u+v)\,dx = \int u\,dx + \int v\,dx$$

Integral de un producto por una constante

$$\int k \cdot u\,dx = k \int u\,dx$$

Integral por partes

$$\int u\,dv = uv - \int v\,du$$

Sustitución simple

$$\int f(g(x)) \cdot g'(x)\,dx = \int f(u)\,du, \quad u = g(x)$$

Distribuciónes comunes

Distribución Binomial

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

Parámetros: n (número de ensayos), p (probabilidad de éxito)

Distribución de Poisson

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Parámetro: λ (tasa de ocurrencia)

Distribución Geométrica

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

Parámetro: p (probabilidad de éxito)

Distribuciones Continuas

Distribución Normal

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

Parámetros: μ (media), σ^2 (varianza)

Distribución Exponencial

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

Parámetro: λ (tasa)

Distribución Uniforme

$$f(x) = \frac{1}{b-a}$$

Parámetros: a (límite inferior), b (límite superior)

Distribución Beta

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

Parámetros: α, β (parámetros de forma)