Capítulo 2

Medidas Estadísticas Básicas

2.1. Medidas estadísticas poblacionales

Sea X una variable aleatoria con función de probabilidad $p(x)$ si es discreta, o función de densidad $f(x)$ si es continua.

Esperanza: La esperanza de una variable aleatoria X es

$$\mu_X = E(X) = \begin{cases} \sum_x x\, p(x) & \text{si } X \text{ es discreta;} \\ \int_{-\infty}^{\infty} x\, f(x)dx & \text{si } X \text{ es continua.} \end{cases}$$

Propiedades: Sean X e Y dos variables aleatorias, y a, b, c constantes. Se verifica:

(i) $E(c) = ...$

(ii) $E(aX + b) = ...$

(iii) $E(X + Y) = ...$

(iv) $E(aX + bY) = ...$

Varianza:

$$\sigma_X^2 = V(X) = E\left[(X - E(X))^2\right].$$

Desviación típica:

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Covarianza:

$$\sigma_{X,Y} = Cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right].$$

Propiedades: Sean X, Y, U, V variables aleatorias y a, b, c, d constantes. Se verifica:

(i) $V(c) = ...$

(ii) $V(aX + b) = ...$

(iii) $V(X + Y) = ...$

(iv) $V(aX + bY) = ...$

(v) $Cov(aX + b, cY + d) = ...$

(vi) $Cov(X + Y, U + V) = ...$

(vii) $Cov(aX + bY, cU + dV) = ...$

Ejemplo 2.1 Sean X_1 y X_2 variables aleatorias incorreladas con media cero y varianzas $\sigma^2_{X_1} = 1$ y $\sigma^2_{X_2} = 2$.

(i) Definimos la nueva variable $Z_1 = 0{,}2\, X_1 + 0{,}7\, X_2$. Calcula $E(Z_1)$ y $V(Z_1)$: ...

(ii) Definimos la nueva variable $Z_2 = -0{,}5\, X_1 + 0{,}1\, X_2$. Calcula $E(Z_2)$, $V(Z_2)$ y $Cov(Z_1, Z_2)$: ...

Coeficiente de correlación lineal:

$$\varrho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Propiedades:

(i) $-1 \leq \varrho_{X,Y} \leq 1$;

(ii) La correlación lineal $\varrho_{X,Y}$ mide la fuerza de la asociación lineal entre X e Y. Si no existe dependencia lineal, $\varrho_{X,Y} = 0$; si la dependencia es lineal directa, $\varrho_{X,Y} > 0$, y si la dependencia es lineal inversa, $\varrho_{X,Y} < 0$.

Vector aleatorio: $\mathbf{X} = (X_1, \ldots, X_p)'$ es un vector cuyas componentes $X_1, \ldots, X_p$ son variables aleatorias.

Esperanza de un vector aleatorio: La esperanza de un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)'$ es el vector compuesto por las esperanzas,

$$\boldsymbol{\mu}_X = E(\mathbf{X}) = (E(X_1), \ldots, E(X_p))'.$$

Matriz de varianzas-covarianzas: La matriz de varianzas-covarianzas de un vector aleatorio X es la matriz definida por:

$$\Sigma_X = V(\mathbf{X}) = E\left[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))'\right].$$

Esta matriz contiene las varianzas en la diagonal, y las covarianzas entre los pares de variables fuera de la diagonal:

$$\Sigma_X = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix},$$

donde

$$\sigma_{i,j} = Cov(X_i, X_j) \quad \text{y} \quad \sigma_i^2 = V(X_i).$$

Matriz de correlaciones: La matriz de correlaciones de un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)$ es

$$P_X = \begin{pmatrix} 1 & \varrho_{12} & \cdots & \varrho_{1p} \\ \varrho_{21} & 1 & \cdots & \varrho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{p1} & \varrho_{p2} & \cdots & 1 \end{pmatrix}$$

donde el elemento (i, j) es

$$\varrho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}.$$

Combinaciones lineales de variables aleatorias:

Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio, y $\mathbf{a} = (a_1, \ldots, a_p)'$, $\mathbf{b} = (b_1, \ldots, b_p)'$ dos vectores de constantes. Definimos dos nuevas variables Z_1 y Z_2 que son combinación lineal de $\mathbf{X}$:

$$Z_1 = \mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots + a_p X_p$$
$$Z_2 = \mathbf{b}'\mathbf{X} = b_1 X_1 + \cdots + b_p X_p$$

Se verifica:

(i) $E(Z_1) = \mathbf{a}'E(\mathbf{X})$;

(ii) $V(Z_1) = \mathbf{a}'V(\mathbf{X})\mathbf{a}$;

(iii) $Cov(Z_1, Z_2) = \mathbf{a}'V(\mathbf{X})\mathbf{b}$.

2.2. Medidas estadísticas muestrales

Media muestral: Sean $\{x_1, x_2, \ldots, x_n\}$ los valores observados de cierta variable X en n individuos. La *media muestral* de X es

$$\bar{x} = ...$$

Varianza muestral: La *varianza muestral* de X es

$$s_X^2 = ...$$

Este valor mide la dispersión de las observaciones alrededor de la media muestral.

Covarianza muestral: Sean $\{x_{11}, x_{21}, \ldots, x_{n1}\}$, $\{x_{12}, x_{22}, \ldots, x_{n2}\}$ los valores observados de las variables X_1 y X_2 en n individuos. Se define la *covarianza muestral* entre X_1 y X_2 como

$$s_{X_1,X_2} = ...$$

La covarianza indica la asociación lineal que existe entre las variables X_1 y X_2. Si la relación entre las variables es lineal directa, la covarianza es positiva. Si la asociación es lineal inversa, entonces la covarianza es negativa.

$$\text{Si } X_2 = X_1, \text{ entonces } s_{X_1,X_2} = s_{X_1}^2.$$

Ejemplo 2.2 Se dispone de los ingresos (X_1) y los gastos (X_2) anuales de 5 individuos. En la tabla de la izquierda están medidos en euros, mientras que en la de la derecha están medidos en miles de euros.

Ind.	Ingresos	Gastos
1	10000	9000
2	7000	8000
3	15000	13000
4	21000	20000
5	14000	13500
Media	13,400	12,700

Ind.	Ingresos	Gastos
1	10	9
2	7	8
3	15	13
4	21	20
5	14	13.5
Media	13.4	12.7

Calcular la covarianza entre los ingresos y los gastos con ambas tablas y comparar el resultado: ...

La covarianza muestral tiene difícil interpretacion, ya que su magnitud depende de las unidades en las que estén medidas las variables. El coeficiente de correlación lineal evita este problema.

Coeficiente de correlación lineal muestral: Se define el *coeficiente de correlación lineal* entre las variables X_1 y X_2 como

$$r_{X_1,X_2} = \dots \qquad .$$

Ejemplo 2.3 Calcular el coeficiente de correlación lineal para las dos tablas de datos del Ejemplo 2.2. ...

Medidas estadísticas multivariantes

Supongamos ahora que se han observado los valores de p variables $X_1, X_2, \ldots, X_p$ en n individuos, de manera que las observaciones se presentan en forma de una matriz $n \times p$,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \tag{2.1}$$

Se define el *vector de medias muestrales* como

$$\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)', \quad \text{donde} \quad \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij},\ j = 1, \ldots, p\,.$$

La *matriz de covarianzas muestral* de X_1, X_2,. . ., X_p es la matriz simétrica formada por las covarianzas entre cada par de variables,

$$S_X = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix},$$

donde s_{ij} se usa para denotar la covarianza s_{X_i,X_j} y s_i^2 para denotar la varianza $s^2_{X_i}$.

La *matriz de correlaciones muestral* de $X_1, X_2, \ldots, X_p$ es

$$R_X = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix},$$

donde r_{ij} denota el coeficiente de correlación lineal r_{X_i,X_j}.

Ejemplo 2.4 Escribir el vector de medias, la matriz de covarianzas y la matriz de correlaciones muestral para los datos del Ejemplo 2.2, tabla derecha.

Ejemplo 2.5 Sean X_1 y X_2 el número de unidades vendidas de dos productos en 25 establecimientos.

X_1	X_2	X_1	X_2	X_1	X_2
191	155	179	158	192	154
195	149	183	147	174	143
181	148	174	150	176	139
183	153	190	159	197	167
176	144	188	151	190	163
208	157	163	137		
189	150	195	155		
197	159	186	153		
188	152	181	145		
192	150	175	140		

El vector de medias y la matriz de covarianzas son

$$\bar{\mathbf{x}} = \begin{pmatrix} 185{,}7 \\ 151{,}1 \end{pmatrix}, \quad S_X = \begin{pmatrix} 95{,}29 & 52{,}87 \\ 52{,}87 & 54{,}36 \end{pmatrix}$$

Sean dos nuevas variables Z_1 y Z_2 que se obtienen a partir de X_1 y X_2 de la forma

$$\begin{aligned} Z_1 &= 0{,}2X_1 + 0{,}7X_2\,, \\ Z_2 &= -0{,}5X_1 + 0{,}1X_2\,. \end{aligned}$$

¿Cómo podemos obtener las medias muestrales de Z_1 y Z_2 a partir de las de X_1 y X_2? ¿y las varianzas muestrales?, ¿y la covarianza muestral entre Z_1 y Z_2?. ...

Momentos muestrales de CLs de variables:

Sean los vectores

$$\mathbf{a} = (a_1, a_2, \ldots, a_p)', \quad \mathbf{X} = (X_1, X_2, \ldots, X_p)'.$$

Definimos una variable Z que es combinación lineal de $\mathbf{X}$:

$$Z = \mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p.$$

Se verifica:

$$\bar{z} = \mathbf{a}'\bar{\mathbf{x}}, \qquad s_Z^2 = \mathbf{a}'S_X\mathbf{a}.$$

Además, si tenemos dos combinaciones lineales de $\mathbf{X}$

$$\begin{aligned} Z_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{21}X_2 + \ldots + a_{p1}X_p\,, \\ Z_2 &= \mathbf{a}_2'\mathbf{X} = a_{12}X_1 + a_{22}X_2 + \ldots + a_{p2}X_p\,. \end{aligned}$$

entonces la covarianza entre ellas es

$$s_{Z_1,Z_2} = \mathbf{a}_1'S_X\mathbf{a}_2.$$

Ejemplo 2.6 A partir de la matrix de covarianzas muestral S_X del Ejemplo 2.5:

(i) Comprueba que los valores y vectores propios son:

Valores	131.52	18.13
Vectores	-0.825	0.565
	-0.565	-0.825

(ii) Sea la variable $Z_1 = -0{,}825\, X_1 - 0{,}565\, X_2$. Calcula la media muestral y la varianza muestral de Z_1.

(iii) Sea la variable $Z_2 = 0{,}565\, X_1 - 0{,}825\, X_2$. Calcula la media muestral y la varianza muestral de Z_2. Calcula también la covarianza muestral entre Z_1 y Z_2.

...