Vectores aleatorios

January 2024

Introducción

El contexto

- `Objetivo`: estudiar k variables sobre una población de individuos (objetos).
- `Algunos ejemplos`:
  - → Las variables meteorológicas como temperatura, humedad y velocidad del viento.
  - → La intensidad y la fase de una señal aleatoria que se miden en los canales de comunicación.
  - → Los parámetros clínicos de los pacientes (como presión arterial, niveles de glucosa, etc.).
- Habitualmente estas variables serán numéricas y tendrán relaciones (correlaciones) entre ellas.
- En algunos casos tendremos variables cualitativas o discretas que nos indicarán grupos de individuos.
- Estas variables se representarán mediante vectores aleatorios sobre un espacio de probabilidad.

Vectores aleatorios

Definiciones

- Un `vector aleatorio` (v.a.) k-dimensional sobre un espacio de probabilidad $(\Omega, \mathcal{S}, \mathcal{P})$ es $\mathbf{X} = (X_1, \ldots, X_k)$ tal que

$$X_i^{-1}(-\infty, x] \in \mathcal{S},$$

para todo $x \in \mathbb{R}$, $i = 1, \ldots, k$.

- `Función de distribución conjunta`, $F : \mathbb{R}^k \to [0, 1]$,

$$F(x_1, \ldots, x_k) := Pr[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k],$$

para todo $x_1, \ldots, x_k \in \mathbb{R}$.

Independencia de variables aleatorias

Definición

- Las variables aleatorias $X_1, \ldots, X_k$ son `independientes` si los sucesos

$$\{X_1 \leq x_1\}, \{X_2 \leq x_2\}, \ldots, \{X_k \leq x_k\}$$

son independientes para todo $x_1, \ldots, x_k \in \mathbb{R}$.

- Esto es equivalente a que

$$F(x_1, \ldots, x_k) = Pr[X_1 \leq x_1] \cdot Pr[X_2 \leq x_2] \cdots Pr[X_k \leq x_k],$$

para todo $x_1, \ldots, x_k \in \mathbb{R}$.

Más definiciones...

Distribuciones marginales

- La función $F_{X_i}(x_i) = Pr[X_i \leq x_i]$ se denomina `función de distribución marginal` i-ésima y corresponde con la función de distribución de la variable aleatoria X_i.
- Las `distribuciones marginales` pueden obtenerse a partir de la distribución conjunta:

$$F_{X_i}(x_i) = F(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty).$$

- Análogamente, la `función de distribución marginal del subvector aleatorio` $(X_{i_1}, \ldots, X_{i_m})$ vendrá dada por

$$\begin{aligned} F_{X_{i_1}, \ldots, X_{i_m}}&(x_{i_1}, \ldots, x_{i_m}) \\ &= F(+\infty, \ldots, +\infty, x_{i_1}, +\infty, \ldots, +\infty, x_{i_m}, +\infty, \ldots, +\infty). \end{aligned}$$

Vector aleatorio absolutamente continuo

- Un vector aleatorio $\mathbf{X}$ es `absolutamente continuo` si existe una función $f : \mathbb{R}^k \to \mathbb{R}$ no negativa (llamada `función de densidad`) tal que

$$F(\mathbf{x}) = F(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(z_1, \ldots, z_k) dz_k \ldots dz_1,$$

para todo $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$.

- Usando el `teorema fundamental del cálculo`, se tiene que en cada punto de continuidad $(x_1, \ldots, x_k)$ de f

$$\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1 \cdots \partial x_k} = f(x_1, \ldots, x_k).$$

- Existen v.a. cuya función de distribución es continua pero que no son absolutamente continuas (tienen una parte singular) y puede ocurrir que $X_1, \dots, X_k$ sean absolutamente continuas y que $(X_1, \dots, X_k)$ no lo sea.
  - → Ejemplo: Si X_1 es una v.a. absolutamente continua, entonces el v.a. $\mathbf{X} = (X_1, X_1)$ es continuo pero no absolutamente continuo.
    - De hecho, es completamente singular ya que está contenido en la recta $y = x$ que tiene medida cero en $\mathbb{R}^2$.
- Esto ocurre si consideramos la notas de unos alumnos y sus medias. En estos casos deberemos eliminar estas variables dependientes del vector.

Vector aleatorio discreto

- Un vector aleatorio $\mathbf{X}$ se dice que es `discreto` si existe un conjunto numerable $S \in \mathbb{R}$ tal que $Pr(X \in S) = 1$.
- `Función masa de probabilidad` de un vector aleatorio discreto:

$$Pr[\mathbf{X} = \mathbf{x}] = Pr[X_1 = x_1, \ldots, X_k = x_k],$$

para todo $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, satisfaciendo:

$\rightarrow$ $Pr[\mathbf{X} = \mathbf{x}] \geq 0, \forall \mathbf{x} \in S$

$\rightarrow$ $\displaystyle\sum_{\mathbf{x} \in S} Pr[\mathbf{X} = \mathbf{x}] = 1$

- `Función de distribución` de un vector aleatorio discreto:

$$F(\mathbf{x}) = Pr[\mathbf{X} \leq \mathbf{x}] = \sum_{\substack{\mathbf{z} \in S \\ \mathbf{z} \leq \mathbf{x}}} Pr[\mathbf{X} = \mathbf{z}],$$

para todo $\mathbf{x} \in \mathbb{R}^k$.

Distribuciones marginales. Caso continuo

Distribución marginal de la variable aleatoria X_i

- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un vector aleatorio continuo con función de densidad f, entonces cada componente X_i es de tipo continuo y su función de distribución es:

$$F_{X_i}(x_i) = Pr[X_i \leq x_i] = \int_{-\infty}^{x_i} f_{X_i}(z_i) dz_i,$$

con

$$f_{X_i}(z_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(z_1, \ldots, z_k) dz_1 \ldots, dz_{i-1} \cdot dz_{i+1} \ldots dz_k,$$

para todo $z_i \in \mathbb{R}$.

- La función de densidad marginal de cualquier subvector se calcularía de igual forma.
- $X_1, \ldots, X_k$ son `independientes` $\Leftrightarrow f(x_1, \ldots, x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k)$.

Distribuciones marginales. Caso discreto

Distribución marginal de la variable aleatoria X_i

- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un vector aleatorio discreto con $Pr[\mathbf{X} \in S] = 1$ y función masa de probabilidad $Pr[\mathbf{X} = \mathbf{x}]$, para tod $\mathbf{x} \in S$.
- Si X_i es una componente arbitraria y por tanto discreta con valores en S_i, entonces su `función masa de probabilidad` puede obtenerse a partir de la conjunta:

$$P[X_i = x_i] = \sum_{\substack{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k \\ (x_1,\ldots,x_i,\ldots,x_n)\in S}} P[X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = x_i, X_{i+1} = x_{i+1}, \ldots, X_k = x_k].$$

- La función masa de probabilidad marginal de cualquier subvector se calcularía de igual forma.
- $X_1, \ldots, X_k$ son `independientes` $\Leftrightarrow$ para todo $(x_1, \ldots, x_k) \in S$,

$$P[X_1 = x_1, \ldots, X_k = x_k] = P[X_1 = x_1] \cdots P[X_k = x_k].$$

Distribuciones condicionadas. Caso continuo

Distribución condicionada al valor de una variable

- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un vector aleatorio continuo con función de densidad f.
- Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R}$ tal que $f_{X_i}(x_i^*) > 0$.
- Se define la `distribución condicionada` de $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$ a $(X_i = x_i^*)$ como la determinada por la función de densidad:

$$
\begin{aligned}
f_{X_1,..,X_{i-1},X_{i+1},...,X_k/X_i=x_i^*}&(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k/x_i^*) \\
&= \frac{f(x_1, \ldots, x_i^*, \ldots, x_k)}{f_{X_i}(x_i^*)}.
\end{aligned}
$$

Distribución condicionada a valores de varias variables

- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un vector aleatorio continuo con función de densidad f.
- Sea $(X_{i_1}, \ldots, X_{i_m})$ un subvector arbitrario y $(x^*_{i_1}, \ldots, x^*_{i_m}) \in \mathbb{R}^m$ tal que:

$$f_{X_{i_1},\ldots,X_{i_m}}(x^*_{i_1}, \ldots, x^*_{i_m}) > 0.$$

- Se define la `distribución condicionada` de $(X_1, \ldots, X_{i_1-1}, X_{i_1+1}, \ldots, X_{i_m-1}, X_{i_m+1}, \ldots, X_k)$ a $(X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m})$ como la determinada por la función de densidad:

$$f_{X_1,\ldots,X_{i_1-1},X_{i_1+1},\ldots,X_{i_m-1},X_{i_m+1},\ldots,X_k/X_{i_1}=x^*_{i_1},\ldots,X_{i_m}=x^*_{i_m}}$$
$$(x_1, \ldots, x_{i_1-1}, x_{i_1+1}, \ldots, x_{i_m-1}, x_{i_m+1}, \ldots, x_k/x^*_i)$$
$$= \frac{f(x_1, \ldots, x^*_{i_1}, \ldots, x^*_{i_m}, \ldots, x_k)}{f_{X_{i_1},\ldots,X_{i_m}}(x^*_{i_1}, \ldots, x^*_{i_m})}.$$

Distribuciones condicionadas. Caso discreto

Distribución condicionada al valor de una variable

- Sea $\mathbf{X} = (X_1, \dots, X_k)$ un vector aleatorio discreto.
- Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R}$ tal que

$$Pr[X_i = x_i^*] > 0.$$

- Se define la `distribución condicionada` de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$ a $(X_i = x_i^*)$ como la determinada por la función masa de probabilidad:

$$Pr[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_k = x_k / X_i = x_i^*]$$
$$= \frac{Pr[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i^*, X_{i+1} = x_{i+1}, \dots, X_k = x_k]}{P[X_i = x_i^*]}$$

para todo $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$ tal que $(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_k) \in S$

Distribución condicionada a valores de varias variables

- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un vector aleatorio discreto.
- Sea $X_{i_1}, \ldots, X_{i_m}$ un subvector arbitrario y $(x^*_{i_1}, \ldots, x^*_{i_m}) \in \mathbb{R}^m$ tal que

$$Pr[X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m}] > 0.$$

- Se define la `distribución condicionada` de $(X_1, \ldots, X_{i_1-1}, X_{i_1+1}, \ldots, X_{i_m-1}, X_{i_m+1}, \ldots, X_k)$ a $(X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m})$ como la determinada por la función masa de probabilidad:

$$Pr[X_1 = x_1, \ldots, X_{i_1-1} = x_{i_1-1}, X_{i_1+1} = x_{i_1+1}, \ldots, X_{i_m-1} = x_{i_m-1},$$

$$X_{i_m+1} = x_{i_m+1}, \ldots, X_k = x_k / X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m}]$$

$$= \frac{Pr[X_1 = x_1, \ldots, X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m}, \ldots, X_k = x_k]}{Pr[X_{i_1} = x^*_{i_1}, \ldots, X_{i_m} = x^*_{i_m}]}$$

para todo $(x_1, \ldots, x_{i_1-1}, x_{i_1+1}, \ldots, x_{i_m-1}, x_{i_m+1}, \ldots, x_k)$, tal que $(x_1, \ldots, x^*_{i_1}, \ldots, x^*_{i_m}, \ldots, x_k) \in S$

Distribución normal multivariante $\mathcal{N}_k(\boldsymbol{\mu}, V)$

Función de densidad

$$f(\mathbf{x}) = \frac{1}{\sqrt{|V|(2\pi)^k}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' V^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

para $\mathbf{x} \in \mathbb{R}^k$, donde $\boldsymbol{\mu}$ es un vector k-dimensional y V es una matriz $k \times k$ simétrica y definida positiva.

Definiciones

- Una matriz simétrica A, de dimensión $k \times k$, se dice que es `definida positiva` si se verifica que $\mathbf{x}' A \mathbf{x} > 0$ para cualquier vector no nulo $\mathbf{x} \in \mathbb{R}^k$.
- Una matriz simétrica A, de dimensión $k \times k$, se dice que es `semidefinida positiva` si se verifica que $\mathbf{x}' A \mathbf{x} \geq 0$ para cualquier vector $\mathbf{x} \in \mathbb{R}^k$.

¿Cómo calcular la inversa de $V = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ con R?

```r
V <- matrix(c(1, 1/2,
              1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
solve(V)
```

```
      [,1] [,2]
[1,]  4/3 -2/3
[2,] -2/3  4/3
```

Normal bivariante

Función de densidad

- Caso bivariante, $k=2$, para $\boldsymbol{\mu}=(0,0)$ y $V=\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$.
- Cálculo de la función de densidad en $\mathbf{x}=(1,1)$ utilizando la función `dmvnorm` de la librería `mvtnorm` de `R`:

▶ Code

```
[1] 0.0943539
```

Función de distribución

- Cálculo (aproximado) de la función de distribución en $\mathbf{x} = (1, 1)$ con la función: `pmvnorm(lower = -Inf, upper = x, mean = mu, sigma = V)`

▶ Code

```
[1] 0.7452036
```

Probabilidad en rectángulos

- Cálculo (aproximado) de las probabilidades en rectángulos dando los límites inferiores y superiores del rectángulo. Por ejemplo, para calcular

$$Pr(-1 < X_1 < 1, -1 < X_2 < 1)$$

usaremos `pmvnorm(lower, upper, mean = mu, sigma = V)`:

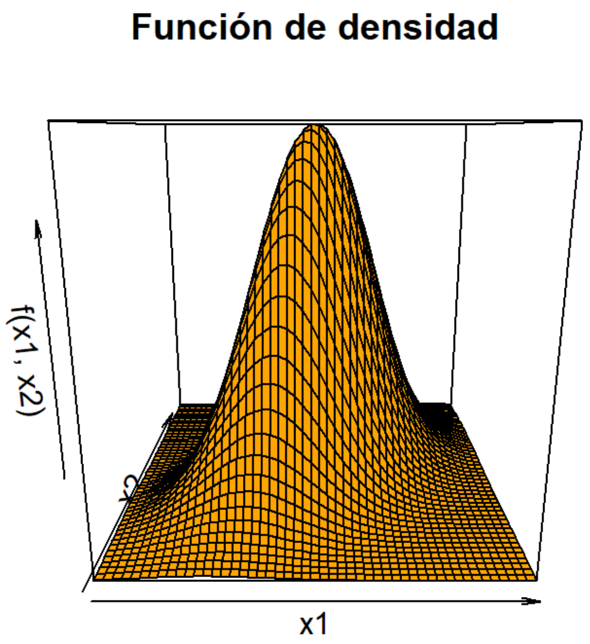
▶ Code

```
[1] 0.4979718
```

¿Cuál es la forma de la densidad?

Su representación gráfica

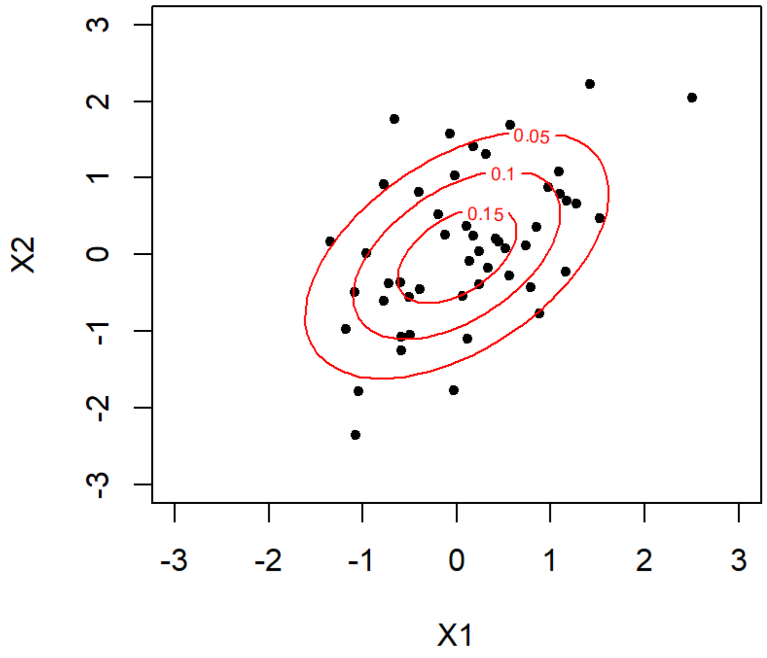
▶ Code



¿Y las curvas de nivel?

Su representación gráfica ($f(x_1, x_2) = c$) y 50 datos simulados de este modelo

▶ Code



Distancia de Mahalanobis

- La distancia de Mahanalobis del vector $\mathbf{x}$ al vector $\boldsymbol{\mu}$ basada en la matriz V:

$$D = \sqrt{(\mathbf{x} - \boldsymbol{\mu})' V^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

- Tiene en cuenta las diferentes escalas de los datos y sus correlaciones.
- Servirá para detectar las observaciones más alejadas del vector de medias que podrían ser observaciones atípicas (`outliers`) que no provengan de nuestra población o contengan errores.
  - $\longrightarrow$ Cuando se pueda, se deberán chequear y, si es posible, corregir o eliminar.
  - $\longrightarrow$ En otros casos, se deberán mantener por ser observaciones correctas que hay que tener en cuenta.

Cálculo de la distancia de Mahalanobis

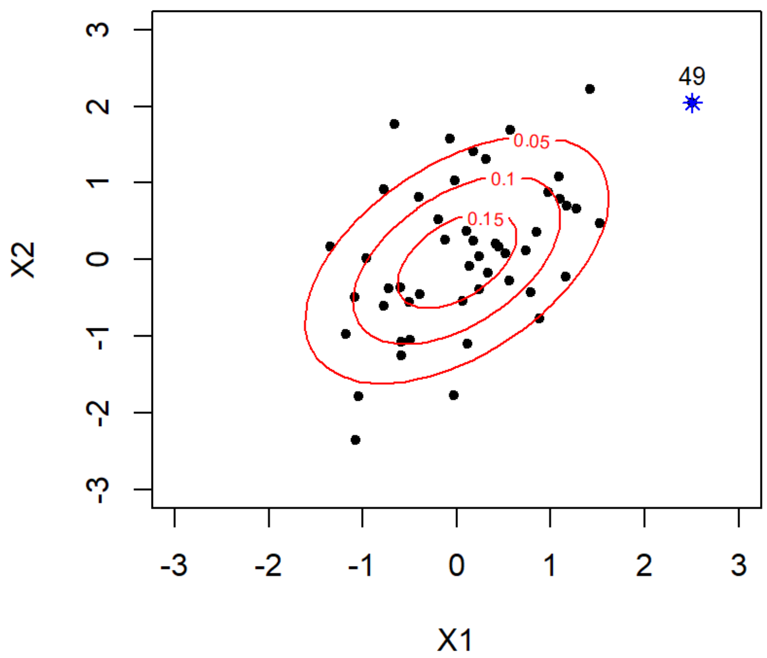
- Para calcular las distancias de Mahalanobis al cuadrado de los datos al vector de medias (teóricas o muestrales) podemos utilizar la función `mahalanobis`.

▶ Code

Distancias de los datos simulados al vector de medias teóricas μ con respecto a V

▶ Code

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
0.05216 0.41016 1.26433 1.66615 2.31591 7.13332 
```



¿Dónde se encuentra la observación más alejada del vector de medias?

▶ Code

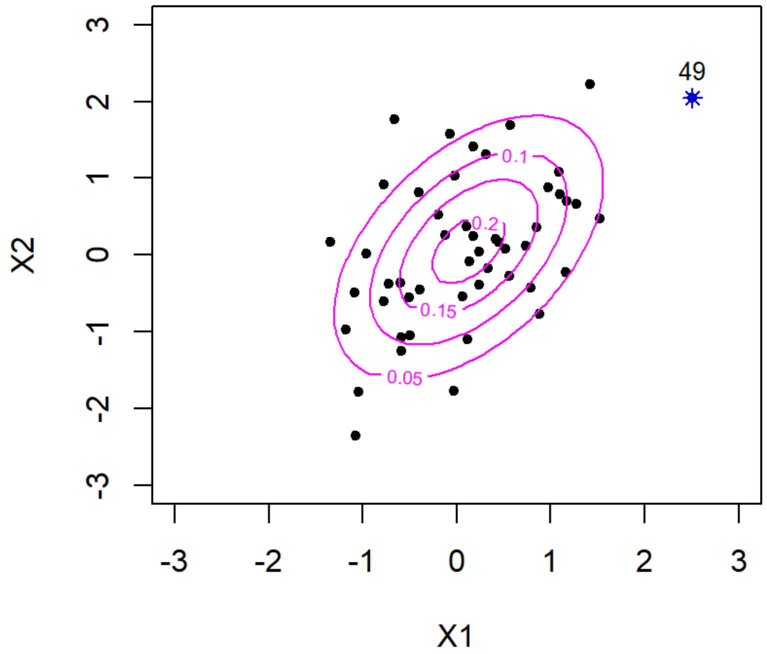
```
[1] 2.509470 2.046512
```

En la práctica las medias y covarianzas serán desconocidas $\longrightarrow$ habrá que estimarlas.

Distancias de los datos simulados al vector de medias muestrales $\overline{\mathbf{x}}$ con respecto a S

▶ Code

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
0.02114 0.67111 1.52636 1.96000 2.64131 8.65906 
```



¿Dónde se encuentra la observación más alejada del vector de medias?

▶ Code

```
[1] 2.509470 2.046512
```

Distribución multinomial $\mathcal{M}_k(n, p_1, \ldots, p_k)$

Modelo multinomial

- $(X_1, \ldots, X_k)$: variables aleatorias que representan el número de veces que ocurre el suceso A_i en un experimento aleatorio repetido n veces con k opciones dadas por $\{A_1, \ldots, A_k\}$ y con probabilidades constantes $p_i = Pr(A_i)$, para $i = 1, \ldots, k$.
- Función masa de probabilidad conjunta:

$$p(x_1, \ldots, x_k) = Pr[X_1 = x_1, \ldots, X_k = x_k] = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

  para enteros no negativos tales que $x_1 + \cdots + x_k = n$, y donde $p_i \in [0, 1]$ satisface $p_1 + \cdots + p_k = 1$.
- Distribuciones marginales:

  $\longrightarrow$ X_i sigue una distribución binomial $B(n, p_i)$, con $E(X_i) = np_i$.

Estadístico de Pearson

Discrepancias entre lo observado y lo esperado

- `Contexto`: Lanzamos un dado $n = 60$ veces, $p_i = 1/6$ para todo i, y los valores esperados son $np_i = 10$, para $i = 1, \ldots, 6$.
- `Objetivo`: Medir las discrepancias entre valores observados y esperados.
- Sea $\mathbf{X} = (X_1, \ldots, X_k)$ un v.a. con distribución multinomial, entonces el estadístico

$$T = \sum_{i=1}^{k} \frac{X_i - np_i}{np_i}$$

sigue una distribución Chi-cuadrado χ^2_{k-1} de Pearson con $k - 1$ grados de libertad, cuando $n \to \infty$.

Medias y covarianzas

Definiciones

- Dado un vector aleatorio,

  → El vector de medias (o esperanza matemática de $\mathbf{X}$) se define como:

  $$\boldsymbol{\mu} := E[\mathbf{X}] = (E[X_1], \ldots, , E[X_k])' = (\mu_1, \ldots, \mu_k)'.$$

  (note que es un vector columna).

  → La matriz de covarianzas (o varianzas-covarianzas) se define como:

  $$V = (\sigma_{i,j}),$$

  donde $\sigma_{i,j}$ es la covarianza entre X_i y X_j, definida como:

  $$\sigma_{i,j} = Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j$$

  Notemos que $\sigma_{i,i} = E[(X_i - \mu_i)^2] = Var(X_i) = \sigma_i^2$.

Cálculo de la esperanza matemática

- La media de cada componente X_i del vector puede calcularse a partir de la distribución conjunta o a partir de la marginal.

  → `Caso discreto`:

$$E[X_i] = \sum_{x_i} x_i Pr[X_i = x_i]$$
$$= \sum_{x_1,\ldots,x_k} x_i Pr[X_1 = x_1, \ldots, X_k = x_k]$$

  → `Caso continuo`:

$$E[X_i] = \int_{\mathbb{R}} x_i f_{X_i}(x_i) dx_i$$
$$= \int_{\mathbb{R}^k} x_i f(x_1, \ldots, x_k) dx_1 \cdots dx_k$$

Esperanza de la transformación $g : \mathbb{R}^k \to \mathbb{R}$. Caso discreto

- Sea $g : \mathbb{R}^k \to \mathbb{R}$ una función medible $\Rightarrow Y = g(\mathbf{X})$ es una variable aleatoria.
- Si $\mathbf{X}$ es de tipo discreto,

$$\exists E[g(\mathbf{X})] \iff \sum_{x_1,\ldots,x_k} |g(x_1,\ldots,x_k)| P[X_1 = x_1,\ldots,X_k = x_k] < \infty$$

- Y en caso de existir:

$$E[g(X_1,..,X_k)] = \sum_{x_1,\ldots,x_k} g(x_1,\ldots,x_k) P[X_1 = x_1,\ldots,X_k = x_k]$$

Esperanza de la transformación $g : \mathbb{R}^k \to \mathbb{R}$. Caso continuo

- Sea $g : \mathbb{R}^k \to \mathbb{R}$ una función medible $\Rightarrow Y = g(\mathbf{X})$ es una variable aleatoria.
- Si $\mathbf{X}$ es de tipo continuo,

$$\exists E[g(\mathbf{X})] \iff \int_{\mathbb{R}^k} |g(x_1, \ldots, x_k)| f_{\mathbf{X}}(x_1, \ldots, x_k) dx_1 \cdots dx_k < \infty$$

- Y en caso de existir:

$$E[g(X_1, \ldots, X_k)] = \int_{\mathbb{R}^k} g(x_1, \ldots, x_k) f_{\mathbf{X}}(x_1, \ldots, x_k) dx_1 \cdots dx_k.$$

Propiedades

- V es una matriz `simétrica` y `semidefinida positiva` (i.e., $\mathbf{x}'V\mathbf{x} \geq \mathbf{0}$, para todo $\mathbf{x} \in \mathbb{R}^k$).
- En forma matricial,

$$V = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

  donde la `esperanza de una matriz aleatoria` se define como la matriz de las esperanzas de cada variable.
- Si X_i y X_j son `independientes`, entonces

$$E[X_i X_j] = E[X_i]E[X_j]$$

  y, por lo tanto, $Cov(X_i, X_j) = \mathbf{0}$. El recíproco no es cierto.
- Si $\mathbf{X} \to \mathcal{N}_k(\boldsymbol{\mu}, V)$, se puede demostrar que $\boldsymbol{\mu}$ es el vector de medias y V es la matriz de covarianzas.

Correlación

- La `correlación (lineal de Pearson)` entre X_i y X_j se define como

$$\rho_{i,j} = Corr(X_i, X_j) = \frac{\sigma_{i,j}}{\sigma_i \sigma_j}$$

  siendo $\rho_{i,i} = Corr(X_i, X_i) = 1$.
- Mide el `grado de relación lineal` entre X_i y X_j.
- Puede demostrarse que

$$-1 \leq \rho_{i,j} \leq 1.$$

- Se dice que X_i y X_j son `incorreladas` si $\rho_{i,j} = 0$.
- Si son independientes serán incorreladas, pero el recíproco no es cierto.
- La `matriz de correlaciones` es $R = (\rho_{i,j})$.

Correlación entre vectores aleatorios

- Análogamente, si $\mathbf{X}$ e $\mathbf{Y}$ son vectores aleatorios (de dimensiones cualesquiera), se define su `matriz de covarianzas` como

$$Cov(\mathbf{X},\mathbf{Y}) = (Cov(X_i, Y_j))$$

y su `matriz de correlaciones` como

$$Corr(\mathbf{X},\mathbf{Y}) = (Corr(X_i, Y_j)).$$

- Puede demostrarse que

$$Cov(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])'].$$

Evidentemente, $Cov(\mathbf{X}) = Cov(\mathbf{X},\mathbf{X})$.

Propiedades

Si $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ son vectores (columna) aleatorios, se verifican las propiedades siguientes:

1. $E[a_1 g_1(\mathbf{X}) + a_2 g_2(\mathbf{X})] = a_1 E[g_1(\mathbf{X})] + a_2 E[g_2(\mathbf{X})]$, donde $a_1, a_2 \in \mathbb{R}$ y g_1 y g_2 son funciones medibles de vectores aleatorios.
2. $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, $E_{\mathbf{X}}[g(\mathbf{Y})] = E_{\mathbf{Y}}[g(\mathbf{Y})]$, donde g es una función medible de un vector aleatorio, $E_{\mathbf{X}}$ denota la esperanza en la distribución conjunta y $E_{\mathbf{Y}}$ en la distribución marginal.
3. Si $\mathbf{X}$ e $\mathbf{Y}$ son independientes, entonces

$$E[g_1(\mathbf{X}) g_2(\mathbf{Y})] = E[g_1(\mathbf{X})] E[g_2(\mathbf{Y})],$$

   donde g_1 y g_2 son funciones medibles cualesquiera de vectores aleatorios.
4. $E[A\mathbf{X} + b] = AE[\mathbf{X}] + b$, $A \in M_{m,k}$, $b' \in \mathbb{R}^m$.
5. $Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$.

6. Si $X_1, \ldots, X_k$ son independientes, $Cov(X_i, X_j) = 0$.
7. $Var(X_i + X_j) = Var(X_i) + 2Cov(X_i, X_j) + Var(X_j)$.
8. $Cov(aX_i + b, cX_j + d) = acCov(X_i, X_j)$, donde $a, b, c, d \in \mathbb{R}$.
9. $Cov(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}'$.
10. $Var(a'\mathbf{X}) = a'Cov(\mathbf{X})a = \sum_{i,j} a_i a_j \sigma_{i,j}$, donde $a \in \mathbb{R}^k$.
11. $Cov(A\mathbf{X} + b) = ACov(\mathbf{X})A'$, donde $A \in M_{m,k}$ y $b' \in \mathbb{R}^m$.
12. Si $X_1, \ldots, X_k$ son independientes, $Corr(X_i, X_j) = 0$.
13. $Corr(aX_i + b, cX_j + d) = Corr(X_i, X_j)$, donde $a, b, c, d \in \mathbb{R}$.
14. $-1 \leq Corr(X_i, X_j) \leq 1$.
15. $Corr(X_i, aX_i + b) = \pm 1$, donde $a, b \in \mathbb{R}$ (según el signo de a).
16. $Corr(\mathbf{X}) = \Delta^{-1} Cov(X) \Delta^{-1}$, donde Δ es la matriz diagonal formada por las desviaciones típicas ($\Delta = diag(\sigma_1, \ldots, \sigma_k)$).
17. $Cov(\mathbf{X}, \mathbf{Y}) = (Cov(X_i, Y_j)) = Cov(\mathbf{Y}, \mathbf{X})'$.

18. $Cov(\mathbf{X} + \mathbf{Z}, \mathbf{Y}) = Cov(\mathbf{X}, \mathbf{Y}) + Cov(\mathbf{Z}, \mathbf{Y})$.
19. Si $\mathbf{X}$ e $\mathbf{Y}$ tienen la misma dimensión, entonces $Cov(\mathbf{X} + \mathbf{Y}) = Cov(\mathbf{X}) + Cov(\mathbf{X}, \mathbf{Y}) + Cov(\mathbf{Y}, \mathbf{X}) + Cov(\mathbf{Y})$.
20. $Cov(A\mathbf{X}, B\mathbf{Y}) = ACov(\mathbf{X}, \mathbf{Y})B'$, donde A y B son matrices (de dimensiones adecuadas).
21. Si $\mathbf{X}$ e $\mathbf{Y}$ independientes, entonces $Cov(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$.

10. $Var(a'\mathbf{X}) = a'Cov(\mathbf{X})a = \sum_{i,j} a_i a_j \sigma_{i,j}$, **donde** $a \in \mathbb{R}^k$

Demostración

- Directamente se tiene que:

$$Var(a'\mathbf{X}) = Cov(a'\mathbf{X}, a'\mathbf{X}) = E[a'(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'a] = a'Cov(\mathbf{X})a$$

- Como consecuencia, se obtiene que la matriz de covarianzas $Cov(\mathbf{X})$ es semidefinida positiva ya que $Var(a'\mathbf{X}) \geq 0$.
- Lo mismo le ocurre a la matriz de correlaciones $Corr(\mathbf{X})$ ya que es la matriz de covarianzas de las v.a. tipificadas $Z_i = (X_i - \mu_i)/\sigma_i$.

Resultados básicos de inferencia

Contexto

- En la práctica, todas las medias, varianzas y covarianzas serán desconocidas por lo que tendremos que estimarlas.
- Para ello dispondremos de una muestra de individuos (objetos) en los que se han medido todas las variables.
- Proporcionamos resultados básicos de inferencia para poder aplicar las técnicas multivariantes que desarrollaremos en temas posteriores.
- Se ilustran estos procedimientos de inferencia con conjuntos de datos de `R`, accesibles con `data()`.

¿Cómo se representan las muestras aleatorias?

Matriz de la muestra aleatoria simple

- En general, nuestra muestra aleatoria se representará como:

i	X_1	X_2	$\ldots$	X_k	Y
$\mathbf{O}_1$	$X_{1,1}$	$X_{1,2}$	$\ldots$	$X_{1,k}$	Y_1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{O}_i$	$X_{i,1}$	$X_{i,2}$	$\ldots$	$X_{i,k}$	Y_i
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{O}_n$	$X_{n,1}$	$X_{n,2}$	$\ldots$	$X_{n,k}$	Y_n

- La variable Y solo se usará para denotar la variable respuesta en regresión.
- En algunos casos usaremos la matriz $M = (X_{i,j})$ que será una matriz aleatoria.

Si no hay grupos supondremos que los objetos $\mathbf{O}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,k})'$ son una m.a.s. de $\mathbf{X}$, es decir, serán vectores (columna) aleatorios independientes con la misma distribución que $\mathbf{X}$.

Si hay grupos supondremos lo mismo en cada grupo.

¿Cómo se muestran los valores muestrales?

Matriz de datos

- En general, nuestra muestra se representará como:

i	x_1	x_2	$\dots$	x_k	y
$\mathbf{o}_1$	$x_{1,1}$	$x_{1,2}$	$\dots$	$x_{1,k}$	y_1
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{o}_i$	$x_{i,1}$	$x_{i,2}$	$\dots$	$x_{i,k}$	y_i
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{o}_n$	$x_{n,1}$	$x_{n,2}$	$\dots$	$x_{n,k}$	y_n

- La variable Y solo se usará para denotar la variable respuesta en regresión.
- En algunos casos usaremos la matriz de datos $M = (x_{i,j})$.

Si no hay grupos, supondremos que los vectores $\mathbf{o}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,k})'$ son una realización de una m.a.s. de $\mathbf{X}$, es decir, serán vectores (columna) con los datos muestrales.

Si hay grupos supondremos lo mismo en cada grupo.

El conjunto de datos `LifeCycleSavings`

Cargamos los datos y visualizamos las primeras filas

▶ Code

```
             sr pop15 pop75     dpi ddpi
Australia 11.43 29.35  2.87 2329.68 2.87
Austria   12.07 23.32  4.41 1507.99 3.93
Belgium   13.17 23.80  4.43 2108.47 3.82
Bolivia    5.75 41.89  1.67  189.13 0.22
Brazil    12.88 42.19  0.83  728.47 4.56
Canada     8.79 31.72  2.85 2982.88 2.43
```

¿Qué información está recogida en el conjunto de datos?

Con la instrucción `help(LifeCycleSavings)` conocemos qué información está contenida en el conjunto:

- `sr`: incremento de los ahorros personales 1960-1970.
- `pop15`: % población menor de 15 años.
- `pop75`: % población mayor de 75.
- `dpi`: ingresos per-capita.

Estimador para el vector de medias $\boldsymbol{\mu}$

- Vector de `medias muestrales`, también llamado `objeto medio`, se define como:

$$\overline{\mathbf{O}} = \overline{\mathbf{X}} = (\overline{X}_1, \ldots, \overline{X}_k)' = \frac{1}{n}\sum_{i=1}^{n} \mathbf{O}_i,$$

donde $\overline{X}_j = \dfrac{1}{n}\sum_{i=1}^{n} X_{i,j}$.

- Se puede demostrar fácilmente que:
  - $\rightarrow$ $E(\overline{\mathbf{O}}) = \boldsymbol{\mu}$ (estimador centrado de $\boldsymbol{\mu}$)
  - $\rightarrow$ $Cov(\overline{\mathbf{O}}) = \dfrac{V}{n}$

¿Dónde se encuentra el vector de medias muestrales?

Propiedad

- $\overline{\mathbf{O}}$ es el punto de $\mathbb{R}^k$ que minimiza la suma de las distancias al cuadrado (`error cuadrático medio`, MSE), es decir, es la solución de

$$\min_{P \in \mathbb{R}^k} MSE = \sum_{i=1}^{n} d^2(\mathbf{O}_i, P),$$

donde d representa la distancia Euclídea, definida para dos vectores $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$ como

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{k} (x_j - y_j)^2}.$$

`MSE`: Mean Square Error

Estimador para la matriz de covarianzas V

- Para estimar $\sigma_{i,j}$ usaremos

  → La `covarianza muestral`: $\hat{\sigma}_{i,j} = \frac{1}{n}\sum_{l=1}^{n}(X_{l,i} - \overline{X}_i)(X_{l,j} - \overline{X}_j)$

  → La `cuasi-covarianza muestral`:

$$S_{i,j} = \frac{1}{n-1}\sum_{l=1}^{n}(X_{l,i} - \overline{X}_i)(X_{l,j} - \overline{X}_j)$$

- Para estimar V usaremos

  → $\hat{V} = (\hat{\sigma}_{i,j}) = \frac{1}{n}\sum_{l=1}^{n}(\mathbf{O}_l - \overline{\mathbf{O}})(\mathbf{O}_l - \overline{\mathbf{O}})'$

  → $S = (S_{i,j}) = \frac{1}{n-1}\sum_{l=1}^{n}(\mathbf{O}_l - \overline{\mathbf{O}})(\mathbf{O}_l - \overline{\mathbf{O}})'$

- Se verifica que:

  → $E(S) = V$ (estimador centrado de V)

Para una distribución normal...

Proposición

Si $\mathbf{X} \to \mathcal{N}_k(\boldsymbol{\mu}, V)$ entonces se verifica que:

- $\overline{O} \to \mathcal{N}_k(\boldsymbol{\mu}, V/n)$
- $\overline{O}$ y $\hat{V}$ son los `estimadores máximo verosímiles` de $\boldsymbol{\mu}$ y V, respectivamente.
- Además, $\overline{O}$ y $\hat{V}$ son `independientes entre sí`. Por tanto, también $\overline{O}$ y S son independientes entre sí.
- La distribución de la matriz aleatoria

$$n\hat{V} = (n-1)S$$

se conoce como `distribución de Wishart`.

Test de normalidad multivariante: Test de Shapiro-Wilk

- Para la aplicación de algunas técnicas multivariantes la hipótesis de normalidad es importante y debe ser contrastada.

$$H_0 : (X_1, \ldots, X_k) \rightarrow \mathcal{N}_k(\boldsymbol{\mu}, V)$$
$$H_1 : (X_1, \ldots, X_k) \nrightarrow \mathcal{N}_k(\boldsymbol{\mu}, V)$$

- Podremos utilizar la función `mshapiro.test` de la librería `mvnormtest` de `R` para realizar el test de normalidad multivariante de Shapiro-Wilk.
  - $\longrightarrow$ Si aplicamos el test a los 50 datos simulados de la normal bivariante lógicamente obtendremos un p-valor que apoya la hipótesis nula.

▶ Code

```
[1] 0.69225
```

Seguimos con `LifeCycleSavings`

Cálculo de las medias muestrales para cada variable

```
mean(datos$sr); mean(datos$pop15); mean(datos$pop75);  mean(datos$dpi); mean(datos$ddpi)
```

```
[1] 9.671
[1] 35.0896
[1] 2.293
[1] 1106.758
[1] 3.7576
```

O bien, podemos calcular todas las características de estas variables

▶ Code

```
       sr             pop15           pop75            dpi         
 Min.   : 0.600   Min.   :21.44   Min.   :0.560   Min.   :  88.94  
 1st Qu.: 6.970   1st Qu.:26.21   1st Qu.:1.125   1st Qu.: 288.21  
 Median :10.510   Median :32.58   Median :2.175   Median : 695.66  
 Mean   : 9.671   Mean   :35.09   Mean   :2.293   Mean   :1106.76  
 3rd Qu.:12.617   3rd Qu.:44.06   3rd Qu.:3.325   3rd Qu.:1795.62  
 Max.   :21.100   Max.   :47.64   Max.   :4.700   Max.   :4001.89  
      ddpi       
 Min.   : 0.220  
 1st Qu.: 2.002  
 Median : 3.000  
 Mean   : 3.758  
 3rd Qu.: 4.478  
 Max.   :16.710  
```

Cálculo de la matriz de covarianzas muestrales

▶ Code

```
          [,1]      [,2]
[1,] 1.1101259 0.8347425
[2,] 0.8347425 1.2075240
```

Cálculo de la matriz de correlaciones muestrales

En este caso es mejor usar correlaciones muestrales que eliminan el efecto de las unidades:

$$R_{i,j} = \frac{S_{i,j}}{S_i S_j},$$

donde $S_i = \sqrt{S_{i,i}}$ y $S_j = \sqrt{S_{j,j}}$

Cálculo de la matriz de correlaciones muestrales

▶ Code

```
              sr       pop15       pop75        dpi        ddpi
sr     1.0000000 -0.45553809  0.31652112  0.2203589  0.30478716
pop15 -0.4555381  1.00000000 -0.90847871 -0.7561881 -0.04782569
pop75  0.3165211 -0.90847871  1.00000000  0.7869995  0.02532138
dpi    0.2203589 -0.75618810  0.78699951  1.0000000 -0.12948552
ddpi   0.3047872 -0.04782569  0.02532138 -0.1294855  1.00000000
```

Observamos que algunas variables tienen correlaciones positivas y otras negativas.