

Material de Nivelación en Probabilidad - Estadística

Mathias Bourel

Instituto de Matemática y Estadística Prof. Rafael Laguardia (IMERL)
Facultad de Ingeniería
Universidad de la República

17 de marzo de 2019

Plan

- 1 Repaso de Probabilidad
 - Espacio de Probabilidad
 - Función de distribución
 - Variables aleatorias discretas y absolutamente continuas
 - Esperanza y Varianza
 - Distribuciones conocidas
- 2 Estimación
- 3 Tests de hipótesis
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos
- 7 Comparación de variables

Espacio Muestral y variable aleatoria

- Sea Ω un espacio muestral, es el conjunto de todos los sucesos elementales de un proceso aleatorio. Por ejemplo:
 - Si lanzo una moneda al aire $\Omega = \{C, X\}$
 - Si tiro un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$
 - Si lanzo un dardo en el segmento $[0, 1]$, $\Omega = [0, 1]$.
- Una σ -álgebra \mathcal{A} sobre Ω es el conjunto de todos los subconjuntos de Ω que serán “probabilizables”.
- Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ es una función que verifica que

$$\forall x \in \mathbb{R} \quad \{w : X(w) \leq x\} \in \mathcal{A}$$

Ejemplos

- 1 si se tiran dos dados, el espacio muestral es

$$\Omega = \{(n_1, n_2) : n_i \in \{1, \dots, 6\}\}$$

y una variable aleatoria podría ser la suma $X(n_1, n_2) = n_1 + n_2$.

- 2 si $\Omega = \{\text{población Montevideo}\}$ una variable aleatoria podría ser la altura $X(w) = \text{altura}$.

Notación

$$[X \leq x] = \{w \in \Omega : X(w) \leq x\}$$

Nociones básicas

- Un evento o suceso A contiene varios sucesos elementales.

- Propiedades de una probabilidad:

- 1 $\mathbb{P}(\emptyset) = 0$

- 2 $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

- 3 Si $A \subset B$ entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$

- 4 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

- 5 $\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n)$ para conjuntos A_n disjuntos dos a dos.

Ejemplo: Lanzamiento de dos dados equilibrados e independientes.

- 1 $\Omega = \{w_1 = (1, 1), w_2 = (1, 2), \dots, w_{36} = (6, 6)\}$

- 2 Si $A =$ la suma de dos valores es igual a 4 entonces

$$\mathbb{P}(A) = \mathbb{P}(\{(1, 3), (2, 2), (3, 1)\}) = \mathbb{P}(1, 1) + \mathbb{P}(2, 2) + \mathbb{P}(3, 1) = \frac{3}{36}$$

En el ejemplo anterior, si definimos como X a la variable aleatoria que devuelve la suma de los valores observados tenemos que $\mathbb{P}(A) = \mathbb{P}(X = 4)$.

Si la variable aleatoria es discreta, la ley de probabilidad de X consiste en dar los valores de las probabilidades $\mathbb{P}(X = x)$ para todo valor posible x de X .

Función de distribución

Si $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad, X una variable aleatoria, la función de distribución de X es $F_X : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$F_X(x) = \mathbb{P}(X \leq x) \quad \forall x \in \mathbb{R}$$

Propiedades:

- 1 $F_X(x) \in [0, 1] \quad \forall x \in \mathbb{R}$
- 2 Si $x_1 \leq x_2$ entonces $F_X(x_1) \leq F_X(x_2)$.
- 3 F_X es continua a la derecha, es decir $\lim_{x \rightarrow a^+} F_X(x) = F_X(a)$.
- 4 $\lim_{x \rightarrow \infty} F_X(x) = 1, \quad \lim_{x \rightarrow -\infty} F_X(x) = 0$.

Función de distribución empírica

Sea X_1, \dots, X_n una sucesión de variables aleatorias independientes con igual distribución F . La función de distribución empírica es

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{(-\infty, t]\}}(X_i) = \frac{1}{n} \#\{\text{cantidad de observaciones} \leq t\}$$

El teorema de Glivenko-Cantelli asegura que con probabilidad 1

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$$

o sea la función de distribución empírica converge uniformemente a la función de distribución F .

Dos tipos de variables aleatorias

Sea $R_X = \{a \in \mathbb{R} : \mathbb{P}(X = a) > 0\}$ el conjunto de los puntos de discontinuidad de F_X .
Se prueba que R_X es numerable.

- X es variable aleatoria discreta si y sólo si $\begin{cases} R_X \text{ es discreto} \\ \mathbb{P}(X \in R_X) = 1 \end{cases}$

Toda variable aleatoria discreta tiene asociada una función de cuantía $p_X : \mathbb{R} \rightarrow \mathbb{R}$ definida por $p_X(x) = \mathbb{P}(X = x) \forall x \in \mathbb{R}$.

Ejemplo: Bernoulli, Binomial, Geométrica, Hipergeométrica, Poisson.

- X es variable aleatoria continua si y sólo si $R_X = \emptyset$.
- X es una variable aleatoria absolutamente continua si y sólo si existe una función $f_X : \mathbb{R} \rightarrow [0, +\infty)$ tal que

$$F_X(x) = \int_{-\infty}^x f_X(s) ds$$

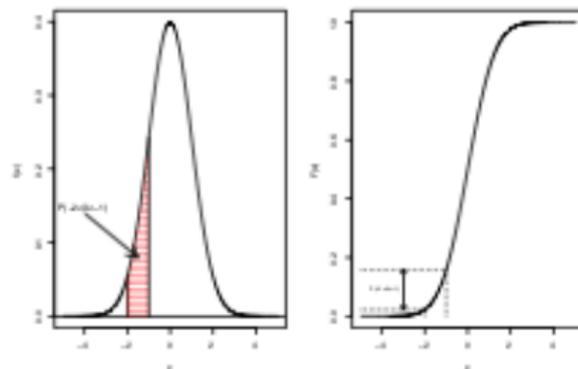
Observar que en este caso:

- $\int_{-\infty}^{+\infty} f_X(s) ds = 1$

- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(s) ds$

Ejemplo: Uniforme, Normal, LogNormal.

Probabilidad y densidad

Figura: $P(-2 < X < 1)$

Esperanza y Varianza

- La *esperanza* de una variable aleatoria X es

$$\mathbb{E}(X) = \sum_{x \in R_X} xp_X(x) \quad \text{si } X \text{ es discreta}$$

$$\mathbb{E}(X) = \int xf_X(x) dx \quad \text{si } X \text{ es absolutamente continua}$$

Un estimador de la esperanza de una variable aleatoria X es

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

donde X_1, \dots, X_n son independientes y todos con la misma distribución que X

- La *varianza* de una variable aleatoria X es

$$\text{Var}(X) = \sigma^2 = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Estimadores de la varianza de una variable aleatoria X son

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

donde X_1, \dots, X_n son independientes y todos con la misma distribución que X

Esperanza y Varianza

1 Para la esperanza

- $\mathbb{E}(X)$ no siempre existe
- Si a es constante entonces $\mathbb{E}(a) = a$
- $\mathbb{E}(X + a) = \mathbb{E}(X) + a$
- $\mathbb{E}(aX) = a\mathbb{E}(X)$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- Si X e Y son independientes entonces $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. El recíproco es falso en general.

2 Para la varianza

- $\text{Var}(X)$ minimiza la función $\mathbb{E}[(X - a)^2]$ ya que $\mathbb{E}[(X - a)^2] = \text{Var}(X) + (\mathbb{E}(X) - a)^2$
- $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- $\text{Var}(X - a) = \text{Var}(X)$
- $\text{Var}(aX) = a^2\text{Var}(X)$
- $\text{Var}(X) = 0 \Leftrightarrow X = a$ c.s
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ donde $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
 $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$
- Si X e Y son independientes entonces $\text{Cov}(X, Y) = 0$ y $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. El recíproco es en general falso.

Variables aleatorias discretas

- ① **Distribución de Bernoulli** $Ber(p)$. Por ejemplo “resultado del lanzamiento de una moneda” que notamos por 1 o 0 según cara o cruz. Si $\mathbb{P}(X = 1) = p$ entonces

- $\mathbb{E}(X) = \sum_{i=0}^1 i\mathbb{P}(X = i) = 0 \times (1 - p) + 1 \times p = p$

- $\mathbb{E}(X^2) = \sum_{i=0}^1 i^2\mathbb{P}(X = i) = 0 \times (1 - p) + 1 \times p = p$

- $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1 - p)$

- ② **Distribución binomial** $B(n, p)$. La variable aleatoria $X \in \{0, 1, \dots, n\}$ cuenta la cantidad de caras obtenidas después de lanzar n veces la moneda, siendo p la probabilidad de obtener cara en cada lanzamiento

- $\mathbb{P}(X = k) = C_k^n p^k (1 - p)^{n-k}$.

- $\mathbb{E}(X) = np$; $\text{Var}(X) = np(1 - p)$

- ③ **Distribución uniforme** $U(\{1, \dots, n\})$. Si la variable aleatoria X toma valores en $\{1, \dots, n\}$

- $\mathbb{P}(X = i) = 1/n$

- $\mathbb{E}(X) = \sum_{i=1}^n i\mathbb{P}(X = i) = \frac{n+1}{2}$

- $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{n^2-1}{12}$

- ④ **Distribución de Poisson** $\mathcal{P}(\lambda)$. $X \in \mathbb{N}$, $\lambda > 0$

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

- $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

Distribución Uniforme

Una variable aleatoria X en $[a, b]$ absolutamente continua tiene distribución uniforme en $[a, b]$ si

$$f_X(x) = \begin{cases} 0 & x \notin [a, b] \\ \frac{1}{b-a} & x \in [a, b] \end{cases}$$

Notación: $X \sim \mathcal{U}[a, b]$.

La función de distribución de $X \sim \mathcal{U}[a, b]$ es:

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

$$\mathbb{E}(X) = \frac{a+b}{2}.$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}.$$

La suma de dos variables aleatorias uniformes no es uniforme.

Distribución Normal

Recordamos:

- 1 la densidad de una normal típica

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

y escribimos $X \sim N(0, 1)$

Distribución Normal

Recordamos:

- 1 la densidad de una normal típica

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

y escribimos $X \sim N(0, 1)$

- 2 la densidad de una normal

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

y escribimos $X \sim N(\mu, \sigma^2)$

Distribución Normal

Recordamos:

- 1 la densidad de una normal típica

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

y escribimos $X \sim N(0, 1)$

- 2 la densidad de una normal

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

y escribimos $X \sim N(\mu, \sigma^2)$

Observaciones:

- Si $X \sim B(n, 0.5)$ y n tiene a infinito, entonces el gráfico de la ley de probabilidad que se obtiene tiende a la curva (simétrica) gaussiana.
- $\mathbb{E}(X) = \mu$ y $\text{Var}(X) = \sigma^2$.
- La gaussiana típica es simétrica al rededor de 0 (alrededor de μ si no es típica). Tiene dos puntos de inflexión en -1 y en 1 ($\pm\sigma$ o $\mu \pm \sigma$ según el caso)
- Si σ es grande el pico de la gaussiana es chico (mucha dispersión a la media) y si σ es chico el pico de la gaussiana es grande (poca dispersión a la media).
- Si $X \sim N(\mu, \sigma^2)$ entonces $z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Distribución Normal

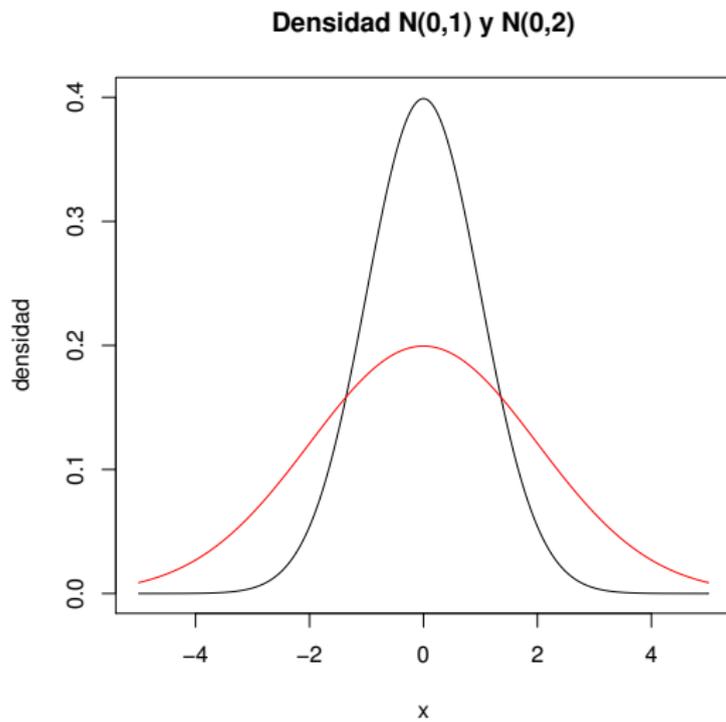
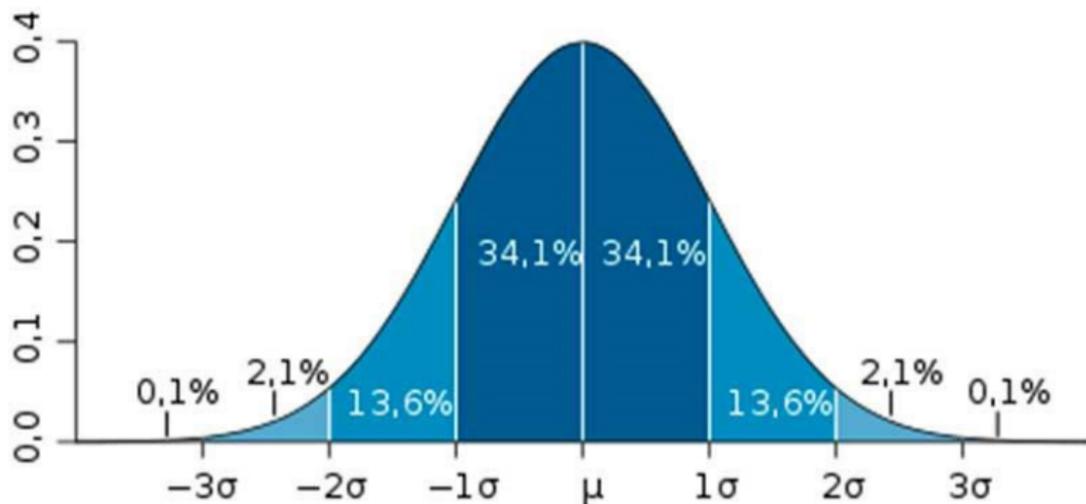


Figura: En negro $\mathcal{N}(0, 2)$ y en rojo $\mathcal{N}(0, 1)$

Distribución Normal



$$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = \mathbb{P}\left(-1 < \frac{X - \mu}{\sigma} < 1\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 2 \times 0,84 - 1 \approx 0,682$$

Esto significa que para una distribución normal, hay un 31,7% de chance de observar un desvío a la media mayor que σ .

Distribución Lognormal

Una variable aleatoria X tiene distribución lognormal si su logaritmo $Y = \ln(X)$ tiene distribución normal $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Su función de densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_Y x} e^{-\frac{(\ln(x)-\mu_Y)^2}{2\sigma_Y^2}} \quad \text{si } x > 0$$

Si X tiene distribución lognormal entonces

$$\mathbb{E}(X) = e^{\mu_Y + \sigma_Y^2/2}$$

$$\text{Var}(X) = (e^{\sigma_Y^2} - 1)e^{2\mu_Y + \sigma_Y^2}$$

En este caso la variable $Z = \frac{\ln(X) - \mu_Y}{\sigma_Y} \sim \mathcal{N}(0, 1)$

Muestra y muestra aleatoria simple - Ley de la media empírica

- En general se hace la hipótesis de que una muestra proviene de realizaciones de una variable aleatoria con una cierta distribución en función de la naturaleza de la característica que se observa.
 - La distribución depende de uno o dos parámetros, en general desconocidos
 - Para ver si una muestra proviene una distribución determinada se puede comparar el histograma obtenido a partir de la muestra con la distribución de la ley. Pero también se puede trazar el QQ-plot, comparar los quantile empíricos con los teóricos, hacer un tesis de hipótesis, etc.
- ② Sea X_1, \dots, X_n una muestra aleatoria simple (X_1, \dots, X_n independientes).
- La media \bar{X}_n es una variable aleatoria
 - Si suponemos que la variable aleatoria es normal con parámetros (μ, σ) entonces la media empírica tiene distribución normal con parámetros $(\mu, \frac{\sigma}{\sqrt{n}})$
 - Si la variable aleatoria es cualquiera, con esperanza μ y varianza σ entonces para valores de n grandes la media empírica tiene distribución normal con parámetros $(\mu, \frac{\sigma}{\sqrt{n}})$ (Teorema Central del Límite).

Plan

- 1 Repaso de Probabilidad
- 2 **Estimación**
 - Estimación Puntual
 - Intervalos de confianza
- 3 Tests de hipótesis
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos
- 7 Comparación de variables

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés.

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- *Muestra*: un subconjunto de tamaño n de la población observada X .

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- *Muestra*: un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- *Muestra*: un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- *Individuo*: un elemento de la población estudiada.

Glosario

- *Población*: conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- *Muestra*: un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- *Individuo*: un elemento de la población estudiada.
- *Parámetro*: valor de interés característico de X o de F_X . Por ejemplo la media, la varianza,...

Glosario

- **Población:** conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- **Muestra:** un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- **Individuo:** un elemento de la población estudiada.
- **Parámetro:** valor de interés característico de X o de F_X . Por ejemplo la media, la varianza,...
- **Variabes:** propiedad común a los individuos de una población. Diferenciamos:
 - **variables cualitativas:** *nominales*, por ejemplo el color del pétalo, el sexo, el color de los ojos, *ordinales*, por ejemplo pequeño, mediano, grande. Se pueden codificar.
 - **variables cuantitativas** (numéricas): por ejemplo el largo del pétalo, el peso, el volumen, etc.
 - *continuas*: $\in \mathbb{R}$. Por ejemplo: temperatura, diámetro de un tronco de árbol, etc.
 - *discretas* o *categorías*: por ejemplo la cantidad de adeptos al fútbol en un grupo de estudiantes.

los valores observados para las variables son los *datos*.

Glosario

- **Población:** conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- **Muestra:** un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- **Individuo:** un elemento de la población estudiada.
- **Parámetro:** valor de interés característico de X o de F_X . Por ejemplo la media, la varianza,...
- **Variables:** propiedad común a los individuos de una población. Diferenciamos:
 - *variables cualitativas: nominales*, por ejemplo el color del pétalo, el sexo, el color de los ojos, o *ordinales*, por ejemplo pequeño, mediano, grande. Se pueden codificar.
 - *variables cuantitativas (numéricas)*: por ejemplo el largo del pétalo, el peso, el volumen, etc.
 - *continuas*: $\in \mathbb{R}$. Por ejemplo: temperatura, diámetro de un tronco de árbol, etc.
 - *discretas o categóricas*: por ejemplo la cantidad de adeptos al fútbol en un grupo de estudiantes.
 los valores observados para las variables son los *datos*.
- **Inferencia Estadística:** proceso mediante al que se llega a conclusiones sobre una población a partir de las observaciones realizadas sobre una muestra de individuos.

Glosario

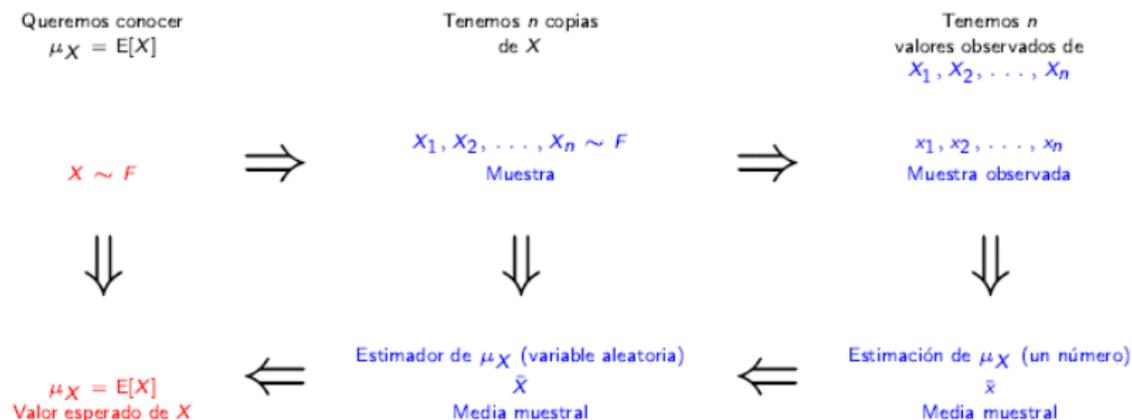
- **Población:** conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- **Muestra:** un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- **Individuo:** un elemento de la población estudiada.
- **Parámetro:** valor de interés característico de X o de F_X . Por ejemplo la media, la varianza,...
- **Variables:** propiedad común a los individuos de una población. Diferenciamos:
 - **variables cualitativas:** *nominales*, por ejemplo el color del pétalo, el sexo, el color de los ojos, *ordinales*, por ejemplo pequeño, mediano, grande. Se pueden codificar.
 - **variables cuantitativas** (numéricas): por ejemplo el largo del pétalo, el peso, el volumen, etc.
 - *continuas:* $\in \mathbb{R}$. Por ejemplo: temperatura, diámetro de un tronco de árbol, etc.
 - *discretas* o *categorías:* por ejemplo la cantidad de adeptos al fútbol en un grupo de estudiantes.
 los valores observados para las variables son los *datos*.
- **Inferencia Estadística:** proceso mediante al que se llega a conclusiones sobre una población a partir de las observaciones realizadas sobre una muestra de individuos.
- **Estadístico:** es una variable aleatoria, función de la muestra aleatoria, $E = E(X_1, X_2, \dots, X_n)$ que se usa para aproximar el valor teórico del parámetro de interés.

Glosario

- **Población:** conjunto de toda la información correspondiente a un valor de interés. La identificamos con una variable aleatoria X y notamos su función de distribución $F_X(x) = \mathbb{P}(X \leq x)$
- **Muestra:** un subconjunto de tamaño n de la población observada X . En general consideramos una familia de n variables aleatorias X_1, \dots, X_n independientes e idénticamente distribuidas (iid) con la distribución de X .
- **Individuo:** un elemento de la población estudiada.
- **Parámetro:** valor de interés característico de X o de F_X . Por ejemplo la media, la varianza,...
- **Variables:** propiedad común a los individuos de una población. Diferenciamos:
 - **variables cualitativas:** *nominales*, por ejemplo el color del pétalo, el sexo, el color de los ojos, *ordinales*, por ejemplo pequeño, mediano, grande. Se pueden codificar.
 - **variables cuantitativas** (numéricas): por ejemplo el largo del pétalo, el peso, el volumen, etc.
 - *continuas:* $\in \mathbb{R}$. Por ejemplo: temperatura, diámetro de un tronco de árbol, etc.
 - *discretas* o *categorías:* por ejemplo la cantidad de adeptos al fútbol en un grupo de estudiantes.
 los valores observados para las variables son los *datos*.
- **Inferencia Estadística:** proceso mediante al que se llega a conclusiones sobre una población a partir de las observaciones realizadas sobre una muestra de individuos.
- **Estadístico:** es una variable aleatoria, función de la muestra aleatoria, $E = E(X_1, X_2, \dots, X_n)$ que se usa para aproximar el valor teórico del parámetro de interés.
- **Estimación:** la evaluación del estimador en una muestra observada x_1, \dots, x_n concreta y que proporciona un valor aproximado del valor del parámetro de interés.

Principio clave: se supone que la muestra es elegida al azar entre todos los individuos posibles para elegir.

Ejemplo



Estimación puntual

| Parámetro desconocido θ_X | Estimador $\hat{\theta}_X = E(X_1, \dots, X_n)$ | Estimación |
|----------------------------------|--|--------------------|
| Media μ_X | media muestral $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ | \bar{x} |
| Proporción p_X | proporción muestral \hat{p}_X | \hat{p}_x |
| Varianza σ_X^2 | varianza muestral $\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ | $\hat{\sigma}_x^2$ |
| Varianza σ_X^2 | varianza muestral corregida $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ | s_x^2 |
| \vdots | \vdots | \vdots |

Bondad de un estimador

Insesgado

El *sesgo* de un estimador es la diferencia entre el valor esperado del estimador y el valor del parámetro considerado $\mathbb{E}(\widehat{\theta}_X) - \theta_X$

| Parámetro desconocido θ_X | Estimador $\widehat{\theta}_X = E(X_1, \dots, X_n)$ | Sesgo |
|----------------------------------|---|--|
| Media μ_X | media muestral $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ | $\mathbb{E}(\bar{X}_n) - \mu_X = 0$ |
| Proporción p_X | proporción muestral \widehat{p}_X | $\mathbb{E}(\widehat{p}_X) - p_X = 0$ |
| Varianza σ_X^2 | varianza muestral $\widehat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ | $\mathbb{E}(\widehat{\sigma}_X^2) - \sigma_X^2 \neq 0$ |
| Varianza σ_X^2 | varianza muestral corregida $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ | $\mathbb{E}(s_X^2) - \sigma_X^2 = 0$ |
| \vdots | \vdots | \vdots |

Bondad de un estimador

Eficiencia

- Un estimador con menos varianza es más eficiente que otro.
Un *estimador eficiente* de θ_X es un estimador insesgado con la menor varianza
- El *error cuadrático medio* de un estimador $\hat{\theta}_X$ de θ_X es

$$ECM(\hat{\theta}_X) = \mathbb{E}(\hat{\theta}_X - \theta_X)^2 = \text{Sesgo}(\hat{\theta}_X)^2 + \text{Var}(\hat{\theta}_X)$$

El error cuadrático medio de un estimador insesgado es igual a su varianza. En efecto:

$$ECM(\hat{\theta}_X) = \mathbb{E}((\hat{\theta}_X - \theta_X)^2) = \mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) + \mathbb{E}(\hat{\theta}_X) - \theta_X \right)^2 \right] \quad (1)$$

$$= \mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right)^2 + 2 \left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right) \left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right) + \left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right)^2 \right] \quad (2)$$

$$= \mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right)^2 \right] + 2\mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right) \left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right) \right] + \mathbb{E} \left[\left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right)^2 \right] \quad (3)$$

$$= \mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right)^2 \right] + 2 \left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right) \underbrace{\mathbb{E}(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X))}_{= \mathbb{E}(\hat{\theta}_X) - \mathbb{E}(\hat{\theta}_X) = 0} + \mathbb{E} \left[\left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right)^2 \right] \quad (4)$$

$$= \mathbb{E} \left[\left(\hat{\theta}_X - \mathbb{E}(\hat{\theta}_X) \right)^2 \right] + \mathbb{E} \left[\left(\mathbb{E}(\hat{\theta}_X) - \theta_X \right)^2 \right] \quad (5)$$

$$= \text{Var}(\hat{\theta}_X) + \text{Sesgo}(\hat{\theta}_X)^2 \quad (6)$$

Bondad de un estimador

Consistencia

La idea es que a medida que el tamaño de la muestra aumenta, aumente también la precisión del estimador.

Bondad de un estimador

Consistencia

La idea es que a medida que el tamaño de la muestra aumenta, aumente también la precisión del estimador.

Decimos que $\hat{\theta}_X$ es un *estimador consistente* de θ_X si cuando $n \rightarrow \infty$ se tiene que

$$\hat{\theta}_X \xrightarrow{c.s} \theta_X$$

es decir $\mathbb{P}(\hat{\theta}_X \xrightarrow[n]{} \theta_X) = 1$.

Intervalo de confianza

Una estimación puntual de un parámetro no toma en cuenta la variabilidad del proceso de estimación debido por ejemplo a :

- el tamaño de la muestra: cuanto más ejemplos, la información debería ser más precisa
- la variabilidad de la población
- conocer otros parámetros de la población.

Veamos ahora un procedimiento que nos permita tener un intervalo real en el cual es probable que pertenezca el valor del parámetro buscado.

Intervalo de confianza

Una estimación puntual de un parámetro no toma en cuenta la variabilidad del proceso de estimación debido por ejemplo a :

- el tamaño de la muestra: cuanto más ejemplos, la información debería ser más precisa
- la variabilidad de la población
- conocer otros parámetros de la población.

Veamos ahora un procedimiento que nos permita tener un intervalo real en el cual es probable que pertenezca el valor del parámetro buscado.

Más precisamente, un estimador por intervalo de confianza de un parámetro θ al nivel $(1 - \alpha) \times 100\%$ es un intervalo $(\hat{a} = a(X_1, \dots, X_n), \hat{b} = b(X_1, \dots, X_n))$ que cumple que

$$\mathbb{P}(\theta \in (\hat{a}, \hat{b})) = 1 - \alpha$$

Un intervalo de confianza para θ con un nivel de confianza $1 - \alpha$ es el valor observado del estimador por intervalo de confianza y con una confianza del $(1 - \alpha) \times 100\%$ el valor del parámetro desconocido de la población está en

$$(\hat{a} = a(x_1, \dots, x_n), \hat{b} = b(x_1, \dots, x_n))$$

Intervalo de confianza

Una estimación puntual de un parámetro no toma en cuenta la variabilidad del proceso de estimación debido por ejemplo a :

- el tamaño de la muestra: cuanto más ejemplos, la información debería ser más precisa
- la variabilidad de la población
- conocer otros parámetros de la población.

Veamos ahora un procedimiento que nos permita tener un intervalo real en el cual es probable que pertenezca el valor del parámetro buscado.

Más precisamente, un estimador por intervalo de confianza de un parámetro θ al nivel $(1 - \alpha) \times 100\%$ es un intervalo $(\hat{a} = a(X_1, \dots, X_n), \hat{b} = b(X_1, \dots, X_n))$ que cumple que

$$\mathbb{P}(\theta \in (\hat{a}, \hat{b})) = 1 - \alpha$$

Un intervalo de confianza para θ con un nivel de confianza $1 - \alpha$ es el valor observado del estimador por intervalo de confianza y con una confianza del $(1 - \alpha) \times 100\%$ el valor del parámetro desconocido de la población está en

$$(\hat{a} = a(x_1, \dots, x_n), \hat{b} = b(x_1, \dots, x_n))$$

Si $\alpha = 0,01$ entonces $(1 - \alpha) \times 100\% = 99\%$.

Si $\alpha = 0,05$ entonces $(1 - \alpha) \times 100\% = 95\%$.

Si $\alpha = 0,1$ entonces $(1 - \alpha) \times 100\% = 90\%$.

Intervalo de confianza: procedimiento general

- 1 Se busca una variable aleatoria Z relacionada con el parámetro desconocido θ y la muestra X_1, \dots, X_n , cuya distribución sea conocida y no dependa del valor del parámetro θ .
- 2 Se plantea la ecuación

$$\mathbb{P}(q_{1-\alpha/2} < Z < q_{\alpha/2}) = 1 - \alpha$$

donde $q_{1-\alpha/2}$ y $q_{\alpha/2}$ son los cuantiles $1 - \alpha/2$ y $\alpha/2$ respectivamente.

- 3 Se resuelve la inecuación en θ para obtener los extremos \hat{a} y \hat{b} .
- 4 Se evalúa para obtener el intervalo de confianza al $(1 - \alpha) \times 100\%$

Ejemplo 1: IdC para media de población normal con varianza conocida

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con σ_X^2 **conocida**.

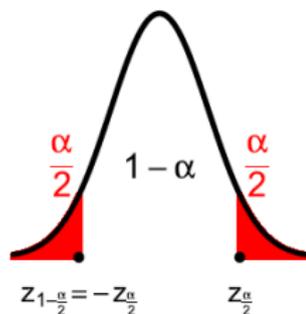
Ejemplo 1: IdC para media de población normal con varianza conocida

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con σ_X^2 **conocida**.

- 1 Se sabe que

$$Z = \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}} \sim \mathcal{N}(0, 1)$$

- 2 La densidad de $Z \sim \mathcal{N}(0, 1)$ es:



Planteo

$$\mathbb{P}(z_{1-\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

$$\mathbb{P}\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2}$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n$$

$$z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n > \mu_X > -z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n$$

$$z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n > \mu_X > -z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \right)$$

Ejemplo 1: IdC para media de población normal con varianza conocida

$$\begin{aligned}
 -z_{\alpha/2} &< \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_{\alpha/2} \\
 -z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} &< \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \\
 -z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n &< -\mu_X < z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n \\
 z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n &> \mu_X > -z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} + \bar{X}_n
 \end{aligned}$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \right)$$

El intervalo de confianza al $(1 - \alpha) \times 100\%$ de μ_X es

$$\left(\bar{x}_n - z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \right)$$

Ejemplo 2: IdC para media de población con muestra grande

Supongamos que X_1, \dots, X_n es una MAS con distribución con parámetros μ_X y σ_X^2 y n es grande.

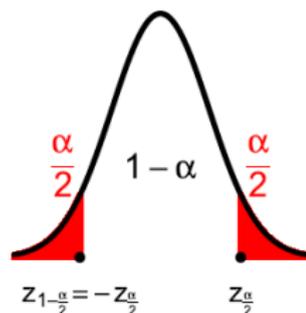
Ejemplo 2: IdC para media de población con muestra grande

Supongamos que X_1, \dots, X_n es una MAS con distribución con parámetros μ_X y σ_X^2 y n es grande.

- 1 Se sabe, basándonos en el Teorema Central del Límite, que

$$Z = \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X / \sqrt{n}} \approx \mathcal{N}(0, 1)$$

- 2 Nos ayudamos de la densidad normal para plantear entonces:



Planteo

$$\mathbb{P}(z_{1-\alpha/2} < Z < z_{\alpha/2}) \approx 1 - \alpha$$

$$\mathbb{P}\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X / \sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Ejemplo 2: IdC para media de población con muestra grande

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2}$$

Ejemplo 2: IdC para media de población con muestra grande

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}$$

Ejemplo 2: IdC para media de población con muestra grande

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n$$

Ejemplo 2: IdC para media de población con muestra grande

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n$$

$$z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n > \mu_X > -z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n$$

Ejemplo 2: IdC para media de población con muestra grande

$$-z_{\alpha/2} < \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}$$

$$-z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n$$

$$z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n > \mu_X > -z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} \right)$$

Ejemplo 2: IdC para media de población con muestra grande

$$\begin{aligned}
 -z_{\alpha/2} &< \frac{\bar{X}_n - \mu_X}{\hat{\sigma}_X/\sqrt{n}} < z_{\alpha/2} \\
 -z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} &< \bar{X}_n - \mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} \\
 -z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n &< -\mu_X < z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} - \bar{X}_n \\
 z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n &> \mu_X > -z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} + \bar{X}_n
 \end{aligned}$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} \right)$$

El intervalo de confianza *aproximado* al $(1 - \alpha) \times 100\%$ de μ_X es

$$\left(\bar{x}_n - z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} \right)$$

Ejemplo 2: IdC para media de población con muestra grande - Aplicación

Supongamos que X_1, \dots, X_n es una MAS con distribución $Ber(p_X)$ con parámetro p_X desconocido y que queremos obtener un intervalo de confianza para μ_X al $(1 - \alpha) \times 100\%$ con $n \geq 30$.

- Del Teorema Central del Límite sabemos que

$$Z = \frac{\hat{p}_X - p_X}{\underbrace{\sqrt{p_X(1-p_X)/n}}_{\sigma_X/\sqrt{n}}} \approx \mathcal{N}(0, 1)$$

También es cierto si tenemos que estimar σ_X :

$$Z = \frac{\hat{p}_X - p_X}{\underbrace{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}}_{\hat{\sigma}_X/\sqrt{n}}} \approx \mathcal{N}(0, 1)$$

- Para muestras grandes, el intervalo de confianza al $(1 - \alpha) \times 100\%$ para p_X es

$$\left(\hat{p}_X - z_{\alpha/2} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{\sqrt{n}}}, \hat{p}_X + z_{\alpha/2} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{\sqrt{n}}} \right)$$

Ejemplo 3: IdC para media de población normal con varianza desconocida

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con σ_X^2 **desconocida**.

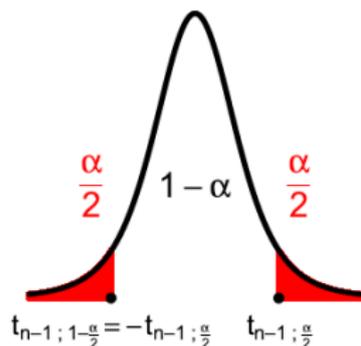
Ejemplo 3: IdC para media de población normal con varianza desconocida

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con σ_X^2 desconocida.

- 1 Se sabe que

$$T = \frac{\bar{X}_n - \mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$$

- 2 La densidad de $T \sim t_{n-1}$ es t de Student con $n - 1$ grados de libertad:



Planteo

$$\mathbb{P}(t_{n-1; 1-\alpha/2} < T < t_{n-1; \alpha/2}) = 1 - \alpha$$

$$\mathbb{P}\left(-t_{n-1; \alpha/2} < \frac{\bar{X}_n - \mu_X}{s_X/\sqrt{n}} < t_{n-1; \alpha/2}\right) = 1 - \alpha$$

Ejemplo 3: IdC para media de población normal con varianza desconocida

$$-t_{n-1;\alpha/2} < \frac{\bar{X}_n - \mu_X}{s_X/\sqrt{n}} < t_{n-1;\alpha/2}$$

Ejemplo 3: IdC para media de población normal con varianza desconocida

$$-t_{n-1;\alpha/2} < \frac{\bar{X}_n - \mu_X}{s_X/\sqrt{n}} < t_{n-1;\alpha/2}$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}}, \bar{X}_n + t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}} \right)$$

Ejemplo 3: IdC para media de población normal con varianza desconocida

$$-t_{n-1;\alpha/2} < \frac{\bar{X}_n - \mu_X}{s_X/\sqrt{n}} < t_{n-1;\alpha/2}$$

El estimador por intervalo de confianza de μ_X es

$$\left(\bar{X}_n - t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}}, \bar{X}_n + t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}} \right)$$

El intervalo de confianza al $(1 - \alpha) \times 100\%$ de μ_X es

$$\left(\bar{x}_n - t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}}, \bar{x}_n + t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}} \right)$$

Distribución chi-cuadrado

- 1 La distribución de la suma de los cuadrados de d variables aleatorias $\mathcal{N}(0, 1)$ independientes es una *chi-cuadrado* con d grados de libertad.
- 2 Si $Z \sim \chi_d^2$ entonces su densidad es

$$f_{\chi_d^2}(t) = \frac{1}{2^{d/2}\Gamma(d/2)} e^{-t/2} t^{d/2-1} \mathbf{1}_{(0,+\infty)}(t)$$

donde $\Gamma(\alpha) = \int_0^{+\infty} e^{-t} t^{\alpha-1} dt$

- 3 $\mathbb{E}(\chi_d^2) = d$
- 4 $\text{Var}(\chi_d^2) = 2d$.

Distribución t de Student

Sean $X \sim \mathcal{N}(0, 1)$ e $Y \sim \chi_k^2$ independientes. La distribución de

$$t_k = \frac{X}{\sqrt{Y/n}}$$

se llama distribución t de Student con k grados de libertad.

Distribución t de Student

Sean $X \sim \mathcal{N}(0, 1)$ e $Y \sim \chi_k^2$ independientes. La distribución de

$$t_k = \frac{X}{\sqrt{Y/n}}$$

se llama *distribución t de Student con k grados de libertad*.

Se verifica que:

- Si $T \sim t_k$ entonces

$$f_T(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}}$$

- $\mathbb{E}(t_k) = 0$
- $\text{Var}(t_k) = \frac{k}{k-2} \quad \forall k > 2$

Distribución chi-cuadrado y t de Student

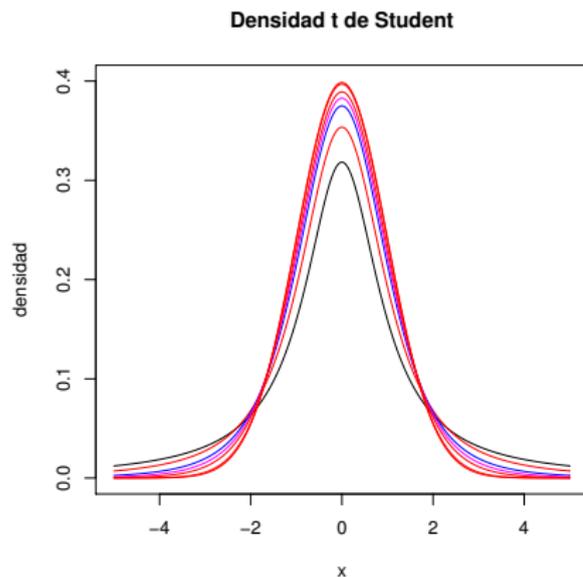


Figura: Densidad t de Student con 1, 4, 6, 10, 50 y 250 grados de libertad.

Distribución chi-cuadrado y t de Student

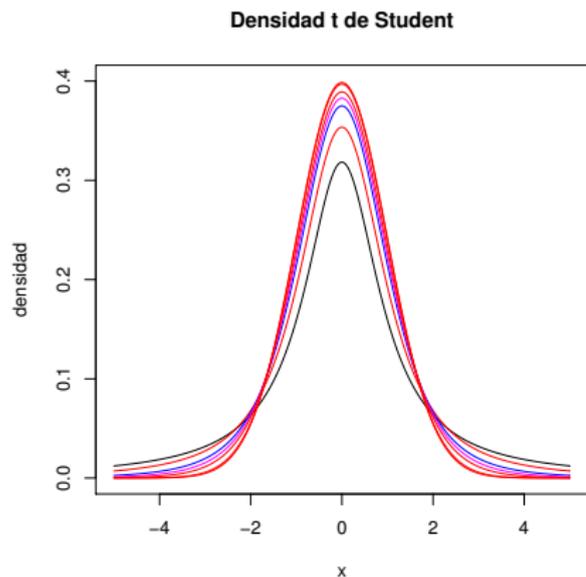


Figura: Densidad t de Student con 1, 4, 6, 10, 50 y 250 grados de libertad.

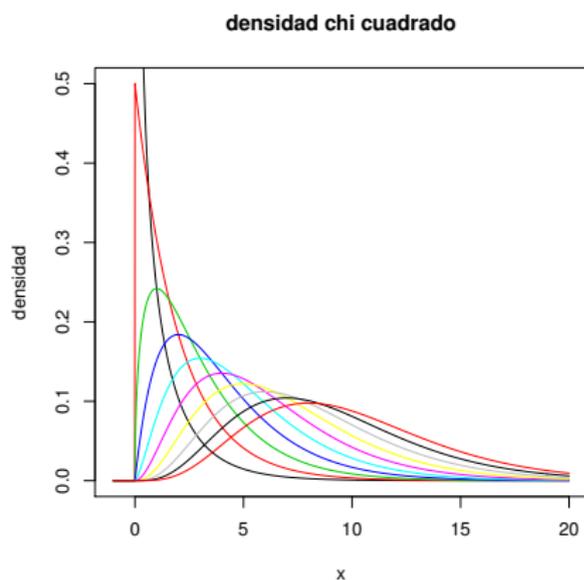


Figura: Densidad χ^2 con 1, 2, \dots , 10 grados de libertad.

Distribución chi-cuadrado y t de Student

¿A qué viene todo esto?

Distribución chi-cuadrado y t de Student

¿A qué viene todo esto?

La varianza corregida reescalada sigue una distribución chi-cuadrado con $n - 1$ grados de libertad. En efecto:

$$\frac{(n-1)s_X^2}{\sigma_X^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma_X^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma_X} \right)^2 \sim \chi_{n-1}^2$$

Distribución chi-cuadrado y t de Student

¿A qué viene todo esto?

La varianza corregida reescalada sigue una distribución chi-cuadrado con $n - 1$ grados de libertad. En efecto:

$$\frac{(n-1)s_X^2}{\sigma_X^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma_X^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma_X} \right)^2 \sim \chi_{n-1}^2$$

El denominador $n - 1$ se debe al hecho siguiente: si conocemos el valor de μ_X tendríamos n variables aleatorias $\frac{X_i - \mu_X}{\sigma_X}$ iid, en cambio, al no conocer el valor de μ_X y estimarlo por \bar{X}_n los grados de libertad son $n - 1$ ya que tenemos $n - 1$ variables aleatorias iid $\frac{X_i - \bar{X}_n}{\sigma_X}$, pues si conocemos $n - 1$ de ellas, se puede deducir fácilmente el valor de la restante.

Ejemplo 4: IdeC para la varianza de una población normal

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con μ_X y σ_X^2 desconocidas.

Ejemplo 4: IdeC para la varianza de una población normal

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con μ_X y σ_X^2 **desconocidas**.

Se quiere hallar un intervalo de confianza al $(1 - \alpha) \times 100\%$ para σ_X^2 .

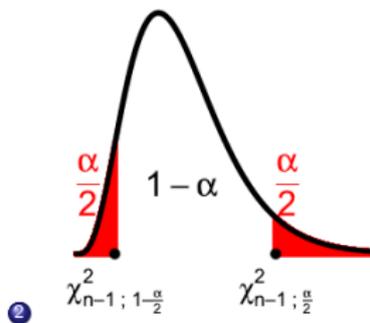
Ejemplo 4: IdeC para la varianza de una población normal

Supongamos que X_1, \dots, X_n es una MAS con distribución normal $\mathcal{N}(\mu_X, \sigma_X^2)$ con μ_X y σ_X^2 desconocidas.

Se quiere hallar un intervalo de confianza al $(1 - \alpha) \times 100\%$ para σ_X^2 .

1 Se sabe que

$$\chi = \frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi_{n-1}^2$$



Planteo

$$\mathbb{P}\left(\chi_{n-1; 1-\alpha/2}^2 < \chi < \chi_{n-1; \alpha/2}^2\right) = 1 - \alpha$$

$$\mathbb{P}\left(\chi_{n-1; 1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1; \alpha/2}^2\right) = 1 - \alpha$$

Ejemplo 4: IdeC para la varianza de una población normal

$$\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1;\alpha/2}^2$$

Ejemplo 4: IdeC para la varianza de una población normal

$$\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1;\alpha/2}^2$$

$$\frac{1}{\chi_{n-1;\alpha/2}^2} > \frac{\sigma_X^2}{(n-1)s_X^2} > \frac{1}{\chi_{n-1;1-\alpha/2}^2}$$

Ejemplo 4: IdeC para la varianza de una población normal

$$\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1;\alpha/2}^2$$

$$\frac{1}{\chi_{n-1;\alpha/2}^2} > \frac{\sigma_X^2}{(n-1)s_X^2} > \frac{1}{\chi_{n-1;1-\alpha/2}^2}$$

$$\frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2} > \sigma_X^2 > \frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}$$

Ejemplo 4: IdeC para la varianza de una población normal

$$\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1;\alpha/2}^2$$

$$\frac{1}{\chi_{n-1;\alpha/2}^2} > \frac{\sigma_X^2}{(n-1)s_X^2} > \frac{1}{\chi_{n-1;1-\alpha/2}^2}$$

$$\frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2} > \sigma_X^2 > \frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}$$

El estimador por intervalo de confianza de σ_X^2 es

$$\left(\frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2} \right)$$

Ejemplo 4: IdeC para la varianza de una población normal

$$\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)s_X^2}{\sigma_X^2} < \chi_{n-1;\alpha/2}^2$$

$$\frac{1}{\chi_{n-1;\alpha/2}^2} > \frac{\sigma_X^2}{(n-1)s_X^2} > \frac{1}{\chi_{n-1;1-\alpha/2}^2}$$

$$\frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2} > \sigma_X^2 > \frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}$$

El estimador por intervalo de confianza de σ_X^2 es

$$\left(\frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2} \right)$$

El intervalo de confianza al $(1-\alpha) \times 100\%$ de σ_X^2 es

$$\left(\frac{(n-1)s_x^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s_x^2}{\chi_{n-1;\alpha/2}^2} \right)$$

Plan

- 1 Repaso de Probabilidad
- 2 Estimación
- 3 Tests de hipótesis**
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos
- 7 Comparación de variables

Introducción

- Se basa en datos muestrales para la toma de una decisión sobre la validez de una conjetura, una *hipótesis*, sobre una población: en general un parámetro como la media μ , o la varianza σ^2 que estamos considerando.
- Esta hipótesis se llama *hipótesis nula* (H_0), la que se supone correcta, y será elegida a menos que la muestra nos indique suficiente evidencia para decir lo contrario (se supone “la inocencia a menos que se pruebe la culpa”).
- Se tiene dos tipos de hipótesis nula:

- 1 Hipótesis simple: del tipo (H_0): $\theta = \theta_0$ como por ejemplo

$$(H_0) : \mu = 5, (H_0) : p = 0,4, (H_0) : \sigma^2 = 4$$

- 2 Hipótesis compuestas: del tipo (H_0): $\theta \geq \theta_0$ o (H_0): $\theta \leq \theta_0$ como por ejemplo

$$(H_0) : \mu \geq 5, (H_0) : p \leq 0,4, (H_0) : \sigma^2 \geq 4$$

Introducción

Si la hipótesis nula no es válida, realizamos el contraste con una *hipótesis alternativa* (H_1):

- Es la hipótesis por la que se inclina el investigador. Se establece en función de lo que se espera descubrir del experimento.
- Se tiene dos tipos de hipótesis alternativa:
 - 1 Hipótesis unilaterales: del tipo $(H_1) : \theta > \theta_0$ o $(H_1) : \theta < \theta_0$ como por ejemplo

$$(H_1) : \mu < 3, (H_1) : p > 0,4, (H_1) : \sigma^2 < 4$$

- 2 Hipótesis bilaterales: del tipo $(H_1) : \theta \neq \theta_0$ como por ejemplo

$$(H_1) : \mu \neq 5, (H_1) : \sigma^2 \neq 4$$

Ejemplo 1

Una empresa produce un tipo de tubos fluorescentes cuya duración media es de por lo menos 1600 unidades de tiempo. Se quiere contrastar esta afirmación a partir de los datos obtenidos en una muestra aleatoria.

- Población: $X =$ "unidades de tiempo de un tubo fluorescente".
- Hipótesis nula (H_0) : $\mu \geq 1600$
- Hipótesis alternativa (H_1) : $\mu < 1600$

Toma de decisión

Una vez que se establecen la hipótesis nula y la hipótesis alternativa y que se dispone de la información proporcionada por la muestra, se toma una decisión decidiendo si se rechaza o no se rechaza la hipótesis nula (H_0).

Toma de decisión

Una vez que se establecen la hipótesis nula y la hipótesis alternativa y que se dispone de la información proporcionada por la muestra, se toma una decisión decidiendo si se rechaza o no se rechaza la hipótesis nula (H_0).

La regla de decisión se basa en el cálculo de un “estadístico”, *el estadístico de prueba*.

Toma de decisión

Una vez que se establecen la hipótesis nula y la hipótesis alternativa y que se dispone de la información proporcionada por la muestra, se toma una decisión decidiendo si se rechaza o no se rechaza la hipótesis nula (H_0).

La regla de decisión se basa en el cálculo de un “estadístico”, *el estadístico de prueba*.

Más allá de la decisión que se tome, la misma puede ser equivocada, porque de lo que se dispone es de una MUESTRA ALEATORIA y nunca se puede tener la certeza de que la hipótesis nula es la correcta.

Toma de decisión

Una vez que se establecen la hipótesis nula y la hipótesis alternativa y que se dispone de la información proporcionada por la muestra, se toma una decisión decidiendo si se rechaza o no se rechaza la hipótesis nula (H_0).

La regla de decisión se basa en el cálculo de un “estadístico”, *el estadístico de prueba*.

Más allá de la decisión que se tome, la misma puede ser equivocada, porque de lo que se dispone es de una MUESTRA ALEATORIA y nunca se puede tener la certeza de que la hipótesis nula es la correcta.

Los tipos de errores que se pueden cometer son dos:

- error de Tipo I: rechazar la hipótesis nula cuando es correcta (importante). La probabilidad de cometer un error de tipo I es α (*nivel de significación*).

$$\alpha = \mathbb{P}(\text{rechazar } (H_0) | (H_0) \text{ es correcta})$$

Toma de decisión

Una vez que se establecen la hipótesis nula y la hipótesis alternativa y que se dispone de la información proporcionada por la muestra, se toma una decisión decidiendo si se rechaza o no se rechaza la hipótesis nula (H_0).

La regla de decisión se basa en el cálculo de un “estadístico”, *el estadístico de prueba*.

Más allá de la decisión que se tome, la misma puede ser equivocada, porque de lo que se dispone es de una MUESTRA ALEATORIA y nunca se puede tener la certeza de que la hipótesis nula es la correcta.

Los tipos de errores que se pueden cometer son dos:

- error de Tipo I: rechazar la hipótesis nula cuando es correcta (importante). La probabilidad de cometer un error de tipo I es α (*nivel de significación*).

$$\alpha = \mathbb{P}(\text{rechazar } (H_0) | (H_0) \text{ es correcta})$$

- error de Tipo II: no rechazar la hipótesis nula cuando es incorrecta. La probabilidad de cometer un error de tipo II es β .

$$\beta = \mathbb{P}(\text{no rechazar } (H_0) | (H_1) \text{ es correcta})$$

La *potencia* π es la probabilidad de rechazar (H_0) cuando es incorrecta

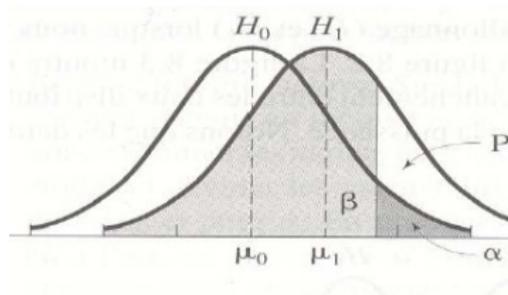
$$\pi = 1 - \beta = \mathbb{P}(\text{rechazar } (H_0) | (H_1) \text{ es correcta})$$

Tabla de contingencia

| Prediction \ Reality | (H_0) correcta | (H_0) incorrecta |
|----------------------|-----------------------------|-----------------------------|
| No rechazar (H_0) | Sin error $1 - \alpha$ | Error de tipo II β |
| Rechazar (H_0) | Error de tipo I α | Sin error $1 - \beta$ |

Los errores de tipo I y II no se pueden cometer simultáneamente, y se prueba que si uno sube, el otro baja, es decir

si $\alpha \uparrow$ entonces $\beta \downarrow$



Estadístico de prueba

- La decisión que se toma se basa en el valor observado del estadístico de prueba. Si los datos muestrales nos proporcionan una evidencia contraria a la hipótesis nula la rechazaremos.
- Para ello necesitamos:
 - la distribución del estadístico del contraste para la muestra, suponiendo cierta la hipótesis nula.
 - el nivel de significación α que definirá la *región de rechazo* o *región crítica* y la *región de aceptación*.
Es el riesgo que uno está dispuesto a correr si se rechaza, erroneamente, la hipótesis nula (H_0).

Estadístico de prueba

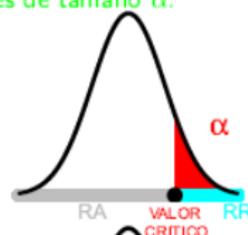
Estadístico del contraste, significación y región de rechazo

Región de rechazo (RR) y de aceptación (RA) en contrastes de tamaño α :

Contraste unilateral derecho $H_1: \theta > \theta_0$



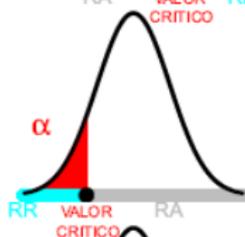
$$RR_\alpha = \{t : t > T_\alpha\} \quad RA_\alpha = \{t : t \leq T_\alpha\}$$



Contraste unilateral izquierdo $H_1: \theta < \theta_0$



$$RR_\alpha = \{t : t < T_{1-\alpha}\} \quad RA_\alpha = \{t : t \geq T_{1-\alpha}\}$$

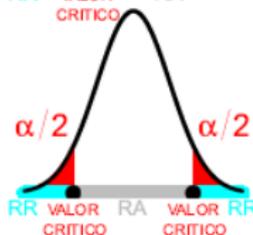


Contraste bilateral $H_1: \theta \neq \theta_0$



$$RR_\alpha = \{t : t < T_{1-\alpha/2} \text{ o } t > T_{\alpha/2}\}$$

$$RA_\alpha = \{t : T_{1-\alpha/2} \leq t \leq T_{\alpha/2}\}$$



Estadístico de prueba

Sea \bar{X}_n una m.a.s. de una población X con media μ y varianza σ^2 , α un nivel de significación, z_α el cuantil α de $N(0,1)$, μ_0 la media de la población bajo H_0 , etc.

| Parámetro | Hipótesis | Estadístico contraste | RR $_{\alpha}$ contraste bilateral |
|-----------|--|--|---|
| Media | Datos normales Varianza conocida | $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ | $\left\{ z : \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha/2} \text{ o } \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \right\}$ |
| | Datos no normales Muestra grande | $\frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} \sim_{ap.} N(0, 1)$ | $\left\{ z : \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} < z_{1-\alpha/2} \text{ o } \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} > z_{\alpha/2} \right\}$ |
| | Datos Bernoulli Muestra grande | $\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/n} \sim_{ap.} N(0, 1)$ | $\left\{ z : \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/n} < z_{1-\alpha/2} \text{ o } \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/n} > z_{\alpha/2} \right\}$ |
| | Datos normales Varianza desconocida | $\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ | $\left\{ t : \frac{\bar{X} - \mu_0}{s/\sqrt{n}} < t_{n-1; 1-\alpha/2} \text{ o } \frac{\bar{X} - \mu_0}{s/\sqrt{n}} > t_{n-1; \alpha/2} \right\}$ |
| Varianza | Datos normales | $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi_{n-1}^2$ | $\left\{ \chi^2 : \frac{(n-1)s^2}{\sigma_0^2} < \chi_{n-1; 1-\alpha/2}^2 \text{ o } \frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1; \alpha/2}^2 \right\}$ |
| Dev. Típ. | Datos normales | $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi_{n-1}^2$ | $\left\{ \chi^2 : \frac{(n-1)s^2}{\sigma_0^2} < \chi_{n-1; 1-\alpha/2}^2 \text{ o } \frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1; \alpha/2}^2 \right\}$ |

Procedimiento general

- Plantear la hipótesis nula y la hipótesis alternativa
- Calcular el valor del estadístico de prueba E y su distribución bajo (H_0) .
- Para un nivel de significación elegido α definir la región crítica \mathcal{R}_α en la que se rechaza (H_0) si el estadístico $E \in \mathcal{R}_\alpha$ y no rechazar en caso contrario.
- Decidir.

El p-valor

Suponiendo que (H_0) es cierta, es la probabilidad de obtener un valor del estadístico de la prueba de hipótesis que sea al menos tan extremo como el observado, o sea es el menor valor de α para el cual se puede rechazar (H_0) .

El p-valor

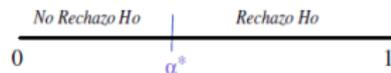
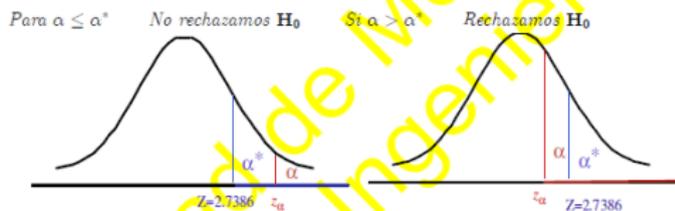
Suponiendo que (H_0) es cierta, es la probabilidad de obtener un valor del estadístico de la prueba de hipótesis que sea al menos tan extremo como el observado, o sea es el menor valor de α para el cual se puede rechazar (H_0).

Regla del p-valor: Una vez que el p-valor ha sido determinado, sabemos que la hipótesis nula se rechaza para cualquier nivel de significación $\alpha \geq \text{p-valor}$; por el contrario, la hipótesis nula no se rechaza cuando $\alpha < \text{p-valor}$. Por lo tanto, el p-valor es un indicador del nivel de admisibilidad de la hipótesis nula: cuanto mayor sea el p-valor, ms confianza podemos tener en la hipótesis nula. El uso del p-valor cambia por completo el enfoque en el contraste de hipótesis. Así, en lugar de fijar a priori el nivel de significación, se calcula el p-valor, que nos permite determinar los niveles de significación para los que se rechaza la hipótesis nula. En resumen, se puede usar en el paso 3) descrito anteriormente como

- Si el p-valor es $< \alpha$, entonces rechazamos (H_0).
- Si el p-valor es $\geq \alpha$, entonces no rechazamos (H_0).

En general, si el p-valor es pequeño tendemos a rechazar (H_0) y si el p-valor es grande tendemos a no rechazar (H_0).

Conociendo el valor α^* , podemos para cualquier α saber la decisión a tomar

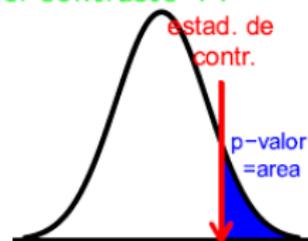


El p-valor

Cálculo del p-valor: t valor observado del estadístico del contraste T :

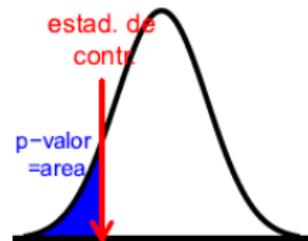
Contraste unilateral derecho $H_1 : \theta > \theta_0$ \Rightarrow

$$p\text{-valor} = P(T \geq t)$$



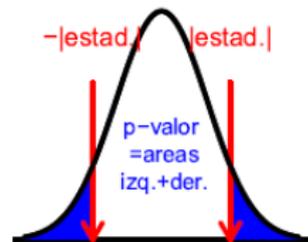
Contraste unilateral izquierdo $H_1 : \theta < \theta_0$ \Rightarrow

$$p\text{-valor} = P(T \leq t)$$



Contraste bilateral $H_1 : \theta \neq \theta_0$ \Rightarrow

$$p\text{-valor} = P(T \leq -|t|) + P(T \geq |t|)$$



Plan

- 1 Repaso de Probabilidad
- 2 Estimación
- 3 Tests de hipótesis
- 4 Tests de normalidad**
 - Test de D'Agostino
 - Test de Shapiro-Wilk
 - Test de Kolmogorov-Smirnov
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos
- 7 Comparación de variables

Test de normalidad de D'Agostino

<http://www.cmat.edu.uy/bioestadistica/archivos/dagostino.pdf>

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Test de normalidad de D'Agostino

<http://www.cmat.edu.uy/bioestadistica/archivos/dagostino.pdf>

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Estadístico de Prueba:

$$E = \sum_{k=1}^{k=n} \frac{(k - \frac{n+1}{2}) X_{(k)}}{\sqrt{n^3(n-1)s_n^2}}$$

donde $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ es la muestra ordenada

Test de normalidad de D'Agostino

<http://www.cmat.edu.uy/bioestadistica/archivos/dagostino.pdf>

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Estadístico de Prueba:

$$E = \sum_{k=1}^{k=n} \frac{(k - \frac{n+1}{2}) X_{(k)}}{\sqrt{n^3(n-1)s_n^2}}$$

donde $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ es la muestra ordenada

Región crítica:

$$E \notin (K_{1,\alpha}(n), K_{2,\alpha}(n))$$

Test de normalidad de D'Agostino

Ejemplo:

| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|-------|
| X_k | 75.02 | 76.42 | 77.94 | 77.74 | 75.22 | 77.2 | 77.74 | 78.32 | 75.62 | 75.58 | 77.6 | 77.5 |
| $X_{(k)}$ | 75.02 | 75.22 | 75.58 | 75.62 | 76.42 | 77.2 | 77.5 | 77.6 | 77.74 | 77.74 | 77.94 | 78.32 |

$$n = 12$$

$$\bar{X}_n = 76,825 \quad s_n^2 = 1,395427 \quad E = \frac{45,17}{\sqrt{12^3(11)(1,395427273)}} = 0,27735$$

$\alpha^* > 0,2 \geq 0,1$ por lo que no se rechaza la hipótesis de que los datos provengan de una distribución normal.

Test de normalidad de Shapiro-Wilk

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Test de normalidad de Shapiro-Wilk

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Estadístico de Prueba:

$$W = \sum_{k=1}^{k=n} \frac{\left(\sum_{i=1}^{[n/2]} a_i (X_{(n-i+1)} - X_{(i)}) \right)^2}{(n-1)s_n^2}$$

donde $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ es la muestra ordenada y los a_i se obtienen de la tabla <http://www.cmat.edu.uy/bioestadistica/archivos/tablas/Shapiro%20Wilk.pdf>

Test de normalidad de Shapiro-Wilk

Sea una muestra X_1, \dots, X_n con distribución F (desconocida)

$$\begin{cases} (H_0) : & F = \text{Normal} \\ (H_1) : & F \neq \text{Normal} \end{cases}$$

Estadístico de Prueba:

$$W = \sum_{k=1}^{k=n} \frac{\left(\sum_{i=1}^{[n/2]} a_i (X_{(n-i+1)} - X_{(i)}) \right)^2}{(n-1)s_n^2}$$

donde $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ es la muestra ordenada y los a_i se obtienen de la tabla <http://www.cmat.edu.uy/bioestadistica/archivos/tablas/Shapiro%20Wilk.pdf>

Región crítica: $W \leq C_\alpha(n)$

Test de Shapiro-Wilk

Ejemplos

① http:

```
//webdelprofesor.ula.ve/ciencias/segninis/Docencia/ANEXO_A_Sahapiro-Wilks.pdf
```

② # Test de normalidad de Shapiro-Wilk

```
set.seed(10)
```

```
x <- rnorm(100) # Creamos una variable normal con 100 valores
```

```
x.test <- shapiro.test(x)
```

```
print(x.test)
```

```
x2 <- runif(100) # Creamos una variable con distribucion uniforme (no normal) con
```

```
x2.test <- shapiro.test(x2)
```

```
print(x2.test)
```

```
Shapiro-Wilk normality test
```

```
data: x
```

```
W = 0.9891, p-value = 0.5911
```

```
Shapiro-Wilk normality test
```

```
data: x2
```

```
W = 0.9285, p-value = 4.082e-05
```

El test de Kolmogorov-Smirnov

Sea una muestra X_1, \dots, X_n con distribución F (desconocida) y sea F_0 una distribución **continua** y conocida (incluido sus parámetros).

$$\begin{cases} (H_0) : & F = F_0 \\ (H_1) : & F \neq F_0 \end{cases}$$

El test de Kolmogorov-Smirnov

Sea una muestra X_1, \dots, X_n con distribución F (desconocida) y sea F_0 una distribución **continua** y conocida (incluido sus parámetros).

$$\begin{cases} (H_0) : & F = F_0 \\ (H_1) : & F \neq F_0 \end{cases}$$

Estadístico de Prueba:

$$E = \sup_{t \in \mathbb{R}} \{|F_n(t) - F_0(t)|\}$$

El test de Kolmogorov-Smirnov

Sea una muestra X_1, \dots, X_n con distribución F (desconocida) y sea F_0 una distribución **continua** y conocida (incluido sus parámetros).

$$\begin{cases} (H_0) : & F = F_0 \\ (H_1) : & F \neq F_0 \end{cases}$$

Estadístico de Prueba:

$$E = \sup_{t \in \mathbb{R}} \{|F_n(t) - F_0(t)|\}$$

Región crítica: $[D_\alpha(n), +\infty)$

Ejemplo Test Kolmogorov-Smirnov

| | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 2.2 | 4.1 | 3.5 | 4.5 | 3.2 | 3.7 | 3.0 | 2.6 |
| 3.4 | 1.6 | 3.1 | 3.3 | 3.8 | 3.1 | 4.7 | 3.7 |
| 2.5 | 4.3 | 3.4 | 3.6 | 2.9 | 3.3 | 3.9 | 3.1 |
| 3.3 | 3.1 | 3.7 | 4.4 | 3.2 | 4.1 | 1.9 | 3.4 |
| 4.7 | 3.8 | 3.2 | 2.6 | 3.9 | 3.0 | 4.2 | 3.5 |

Probar que los datos si se ajustan a una distribución normal con $\mu=3.5$ y $\sigma=0.7$

| OBSERVACION | F.REL.ACM | F(X) | $ H_{i-1} - F_i $ | $ H_i - F_i $ |
|-------------|-----------|--------|-------------------|-----------------|
| 1.60000 | 0.025000 | 0.0033 | 0.0033 | 0.0217 |
| 1.90000 | 0.050000 | 0.0111 | 0.0139 | 0.0389 |
| 2.20000 | 0.075000 | 0.0317 | 0.0183 | 0.0433 |
| 2.50000 | 0.100000 | 0.0766 | 0.0016 | 0.0234 |
| 2.60000 | 0.150000 | 0.0993 | 0.0007 | 0.0507 |
| 2.90000 | 0.175000 | 0.1957 | 0.0457 | 0.0207 |
| 3.00000 | 0.225000 | 0.2375 | 0.0625 | 0.0125 |
| 3.10000 | 0.325000 | 0.2839 | 0.0589 | 0.0411 |
| 3.20000 | 0.400000 | 0.3341 | 0.0091 | 0.0659 |
| 3.30000 | 0.475000 | 0.3875 | 0.0125 | 0.0875 |
| 3.40000 | 0.550000 | 0.4432 | 0.0318 | 0.1068 * |
| 3.50000 | 0.600000 | 0.5000 | 0.0500 | 0.1000 |
| 3.60000 | 0.625000 | 0.5568 | 0.0432 | 0.0682 |
| 3.70000 | 0.700000 | 0.6125 | 0.0125 | 0.0875 |
| 3.80000 | 0.750000 | 0.6659 | 0.0341 | 0.0841 |
| 3.90000 | 0.800000 | 0.7161 | 0.0339 | 0.0839 |
| 4.10000 | 0.850000 | 0.8043 | 0.0043 | 0.0457 |
| 4.20000 | 0.875000 | 0.8414 | 0.0086 | 0.0336 |
| 4.30000 | 0.900000 | 0.8735 | 0.0015 | 0.0265 |
| 4.40000 | 0.925000 | 0.9007 | 0.0007 | 0.0243 |
| 4.50000 | 0.950000 | 0.9234 | 0.0016 | 0.0266 |
| 4.70000 | 1.000000 | 0.9568 | 0.0068 | 0.0432 |

Ejemplo Test Kolmogorov-Smirnov

$D_e = 0.1068$, Es el mayor valor de las dos últimas columnas.

$D_T = 0.2150$, Para un nivel de significancia $\alpha = 0.05$ y $n=40$.

Como $D_e < D_T$, no se rechaza la hipótesis nula de que los datos se ajustan a una distribución normal con $\mu = 3.5$ y $\sigma = 0.7$

Plan

- 1 Repaso de Probabilidad
- 2 Estimación
- 3 Tests de hipótesis
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal**
 - Matrices y vectores
 - Proyección ortogonal
- 6 Matriz de datos
- 7 Comparación de variables

Repaso Álgebra Lineal: matrices y vectores

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n \times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j .

Repaso Álgebra Lineal: matrices y vectores

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n \times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j .

- La traspuesta de una matriz $X \in \mathcal{M}_{n \times p}$ es la matriz $X' \in \mathcal{M}_{p \times n}$ definida por $x'_{ij} = x_{ji}$ para todo i, j .

Por ejemplo, si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ entonces $X' = \begin{pmatrix} 2 & 1 & 6 \\ -1 & 2 & 3 \\ 3 & -3 & 2 \\ 0 & 2 & 10 \end{pmatrix}$

Repaso Algebra Lineal: matrices y vectores

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n \times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j .

- La traspuesta de una matriz $X \in \mathcal{M}_{n \times p}$ es la matriz $X' \in \mathcal{M}_{p \times n}$ definida por $x'_{ij} = x_{ji}$ para todo i, j .

Por ejemplo, si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ entonces $X' = \begin{pmatrix} 2 & 1 & 6 \\ -1 & 2 & 3 \\ 3 & -3 & 2 \\ 0 & 2 & 10 \end{pmatrix}$

- Un vector columna \mathbf{x} es una matriz con una única columna y un vector fila \mathbf{x}' es una matriz con una sola fila. Si $\mathbf{x} = \begin{pmatrix} 1 \\ 4 \\ -3 \end{pmatrix}$ entonces $\mathbf{x}' = (1 \quad 4 \quad -3)$.

Repaso Álgebra Lineal: operaciones

- **Suma de matrices.** Si X e Y son dos matrices $n \times p$, entonces sumamos entrada a entrada, obteniendo una nueva matriz de tamaño $n \times p$. Por ejemplo, si

$$X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix} \text{ e } Y = \begin{pmatrix} 4 & 2 & 5 & -1 \\ 8 & 1 & 3 & 4 \\ 2 & -3 & 1 & -3 \end{pmatrix} \text{ entonces}$$

$$X + Y = \begin{pmatrix} 6 & 1 & 8 & -1 \\ 9 & 3 & 0 & 6 \\ 8 & 0 & -1 & 7 \end{pmatrix}$$

- **Multiplicación por un escalar.** Si X es de tamaño $n \times p$ y $\lambda \in \mathbb{R}$ entonces la matriz λX es la matriz que se obtiene de X multiplicando todas sus entradas por λ . Por ejemplo si

$$X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix} \text{ entonces}$$

$$3X = \begin{pmatrix} 6 & -3 & 9 & 0 \\ 3 & 6 & -9 & 6 \\ 18 & 9 & -6 & 30 \end{pmatrix}$$

Multiplicación de matrices

Si $X = ((x_{ij})) \in \mathcal{M}_{n \times p}$ e $Y = ((y_{ij})) \in \mathcal{M}_{p \times m}$ entonces la matriz $Z = XY \in \mathcal{M}_{n \times m}$ y la entrada z_{ij} de Z es

$$z_{ij} = \sum_{k=1}^p x_{ik} y_{kj}$$

Por ejemplo si $X = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}$ e $Y = \begin{pmatrix} -1 & 2 & -2 \\ 3 & 2 & -4 \end{pmatrix}$ entonces

$$Z = XY = \begin{pmatrix} 7 & 10 & -14 \\ 13 & 6 & -14 \end{pmatrix}$$

Cuidado con el orden:

- Observar que en el ejemplo anterior no se puede hacer YX ...
- Si $\mathbf{u} \in \mathbb{R}^n$ es un vector entonces $\mathbf{u}'\mathbf{u} \in \mathbb{R}$ pero $\mathbf{u}\mathbf{u}' \in \mathcal{M}_{n \times n}$...

La traspuesta de una matriz verifica:

- $(X + Y)' = X' + Y'$
- $(aX)' = aX'$
- $(XY)' = Y'X'$

Interpretación geométrica de los vectores

Un vector fila $\mathbf{x}' = (x_1 \quad x_2 \quad \dots \quad x_n)$ o un vector columna $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ es un punto en \mathbb{R}^n .

El *producto escalar* entre dos vectores \mathbf{x} e \mathbf{y} de \mathbb{R}^n se nota por

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}' \mathbf{y}$$

Se dice que dos vectores son *ortogonales* si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

La *norma* del vector \mathbf{x} es

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Es posible probar que

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\mathbf{x}, \mathbf{y})$$

Proyección ortogonal sobre un vector no nulo

Sea \mathbf{u} un vector no nulo. Entonces la *proyección ortogonal* de \mathbf{x} sobre \mathbf{u} es el vector

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \mathbf{u} = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\|\mathbf{u}\|} \frac{\mathbf{u}}{\|\mathbf{u}\|}$$

Observar que si \mathbf{u} es unitario (tiene norma 1) entonces

$$P_{\mathbf{u}}(\mathbf{x}) = (\mathbf{x}' \mathbf{u}) \mathbf{u}$$

La coordenada de $P_{\mathbf{u}}(\mathbf{x})$ es $\frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\|\mathbf{u}\|}$.

Por ejemplo si $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ entonces la proyección ortogonal de \mathbf{x} sobre \mathbf{u} es:

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{1}{5} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

y la coordenada de $P_{\mathbf{u}}(\mathbf{x})$ es $\frac{1}{\sqrt{5}}$.

Si $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ es una base ortogonal de S , entonces la proyección ortogonal de \mathbf{x} sobre S es el vector

$$P_S(\mathbf{x}) = \frac{\mathbf{x}' \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{x}' \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 + \dots + \frac{\mathbf{x}' \mathbf{u}_r}{\|\mathbf{u}_r\|^2} \mathbf{u}_r$$

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$. Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si existe una matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$. Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si existe una matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.
Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$. Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si existe una

matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.

Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Una matriz $U \in \mathcal{M}_{n \times n}$ es *ortogonal* si U es invertible y $U'U = UU' = I_n$.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$. Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si existe una

matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.

Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Una matriz $U \in \mathcal{M}_{n \times n}$ es *ortogonal* si U es invertible y $U'U = UU' = I_n$.

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector \mathbf{x} se le llama *vector propio* asociado al valor propio λ .

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$. Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si existe una

matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.

Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Una matriz $U \in \mathcal{M}_{n \times n}$ es *ortogonal* si U es invertible y $U'U = UU' = I_n$.

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

- 1 Si $A \in \mathcal{M}_{n \times n}$ es una matriz cualquiera entonces

$$\langle Ax, y \rangle = \langle x, A'y \rangle$$

Esto es porque $\langle x, y \rangle = x'y$

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

- 1 Si $A \in \mathcal{M}_{n \times n}$ es una matriz cualquiera entonces

$$\langle Ax, y \rangle = \langle x, A'y \rangle$$

Esto es porque $\langle x, y \rangle = x'y$

- 2 Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces

$$\langle Ax, y \rangle = \langle x, Ay \rangle$$

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

- 1 Si $A \in \mathcal{M}_{n \times n}$ es una matriz cualquiera entonces

$$\langle Ax, y \rangle = \langle x, A'y \rangle$$

Esto es porque $\langle x, y \rangle = x'y$

- 2 Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces

$$\langle Ax, y \rangle = \langle x, Ay \rangle$$

- 3 Sea A simétrica ahora y consideramos u_1 vector propio de A asociado a λ_1 y u_2 vector propio de A asociado a λ_2 entonces

$$\langle Au_1, u_2 \rangle = \langle \lambda_1 u_1, u_2 \rangle = \lambda_1 \langle u_1, u_2 \rangle$$

que es igual a

$$\langle u_1, Au_2 \rangle = \langle u_1, \lambda_2 u_2 \rangle = \lambda_2 \langle u_1, u_2 \rangle$$

Por lo tanto

$$(\lambda_1 - \lambda_2) \langle u_1, u_2 \rangle = 0$$

y al ser $\lambda_1 \neq \lambda_2$ se tiene que $u_1 \perp u_2$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n \times n}$ y $A'A \in \mathcal{M}_{p \times p}$ son simétricas.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n \times n}$ y $A'A \in \mathcal{M}_{p \times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'A x = 0$ y entonces

$$0 = \langle A'A x, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n \times n}$ y $A'A \in \mathcal{M}_{p \times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n \times n}$ y $A'A \in \mathcal{M}_{p \times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'A x = 0$ y entonces

$$0 = \langle A'A x, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

- Los valores propios de $A'A$ son no negativos.

En efecto si λ es valor propio $\langle A'A x, x \rangle = \lambda \|x\|^2$ por un lado, y $\langle A'A x, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0$ por otro. Entonces

$$\lambda \|x\|^2 = \|Ax\|^2 \geq 0 \Rightarrow \lambda = \frac{\|Ax\|^2}{\|x\|^2}$$

ya que $\|x\|^2 \neq 0$ por ser x vector propio y por lo tanto $\lambda \geq 0$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n \times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n \times n}$ y $A'A \in \mathcal{M}_{p \times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

- Los valores propios de $A'A$ son no negativos.

En efecto si λ es valor propio $\langle A'Ax, x \rangle = \lambda \|x\|^2$ por un lado, y $\langle A'Ax, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0$ por otro. Entonces

$$\lambda \|x\|^2 = \|Ax\|^2 \geq 0 \Rightarrow \lambda = \frac{\|Ax\|^2}{\|x\|^2}$$

ya que $\|x\|^2 \neq 0$ por ser x vector propio y por lo tanto $\lambda \geq 0$.

Descomposición espectral

Teorema Espectral para matrices simétricas:

Sea $A \in \mathcal{M}_{n \times n}$ simétrica. Entonces A es diagonalizable en una base ortonormal de \mathbb{R}^n y podemos escribir:

$$A = UDU' \quad \text{con } U \in \mathcal{M}_{n \times n} \text{ ortogonal}$$

$$\begin{aligned} A &= (u_1 \ u_2 \ \dots \ u_n) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ u_n' \end{pmatrix} \\ &= (u_1 \ u_2 \ \dots \ u_n) \begin{pmatrix} \lambda_1 u_1' \\ \lambda_2 u_2' \\ \vdots \\ \lambda_n u_n' \end{pmatrix} = \sum_{i=1}^n \lambda_i u_i u_i' \end{aligned}$$

La descomposición matricial $A = \sum_{i=1}^n \lambda_i u_i u_i'$ en suma de n matrices se llama *descomposición espectral* de A .

Importante: no confundir con $\sum_{i=1}^n \lambda_i u_i' u_i = \sum_{i=1}^n \lambda_i = \lambda_1 + \dots + \lambda_n \in \mathbb{R}$

Descomposición en valores singulares

Si $A \in \mathcal{M}_{n \times p}$ con $n \geq p$ tiene rango r entonces A se puede descomponer como

$$A = UDV'$$

con

- $V \in \mathcal{M}_{p \times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de $A'A$** .
- $D \in \mathcal{M}_{r \times r} = \text{diag}(\sigma_1, \dots, \sigma_r)$, donde $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ donde los r valores singulares de A corresponden a la raíz cuadrada de los valores propios no nulos de $A'A$ ($\sigma_j = \sqrt{\lambda_j} \forall j = 1, \dots, r$), en su diagonal principal ordenados de mayor a menor.
- $U \in \mathcal{M}_{n \times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de AA'** .

Entonces la *descomposición en valores singulares* de A es

$$A = \sum_{i=1}^r \sigma_i u_i v_i'$$

Descomposición en valores singulares

Si $A \in \mathcal{M}_{n \times p}$ con $n \geq p$ tiene rango r entonces A se puede descomponer como

$$A = UDV'$$

con

- $V \in \mathcal{M}_{p \times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de $A'A$** .
- $D \in \mathcal{M}_{r \times r} = \text{diag}(\sigma_1, \dots, \sigma_r)$, donde $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ donde los r valores singulares de A corresponden a la raíz cuadrada de los valores propios no nulos de $A'A$ ($\sigma_j = \sqrt{\lambda_j} \forall j = 1, \dots, r$), en su diagonal principal ordenados de mayor a menor.
- $U \in \mathcal{M}_{n \times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de AA'** .

Entonces la *descomposición en valores singulares* de A es

$$A = \sum_{i=1}^r \sigma_i u_i v_i'$$

Obs: todo se basa en que $A'A$ y AA' tienen los mismos valores propios no nulos y que si v es vector propio de $A'A$ entonces Av es vector propio de AA' asociado al mismo valor propio. En efecto, por construcción, si λ_j es valor propio de $A'A$ asociado a w_j entonces λ_j es valor propio de AA' asociado a u_j ya que

$$AA'(u_j) = AA' \left(\frac{Av_j}{\sigma_j} \right) = A \left(\frac{\lambda_j v_j}{\sigma_j} \right) = \lambda_j u_j$$

Descomposición SVD - Ejemplo

Ejemplo: Hallar la descomposición SVD de $A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$

1 La matriz $A'A$ es $\begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$ con valores y subespacios propios

$$\left\{ \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\} \leftrightarrow 0, \quad \left\{ \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} \right\} \leftrightarrow 9, \quad \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\} \leftrightarrow 25$$

2 • $D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \quad V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} \\ -1/\sqrt{2} & -1/\sqrt{18} \\ 0 & 4/\sqrt{18} \end{pmatrix}$

• Haciendo $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \times \frac{1}{5} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$,

$$\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix} \times \frac{1}{3} = \begin{pmatrix} \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{pmatrix}.$$

La matriz U es $U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$

• Por lo tanto

$$A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{pmatrix} = UDV'$$

Descomposición SVD

Si $n \geq p$, entonces podemos considerar $A' \in \mathcal{M}_{p \times n}$ y $n \geq p$ (como en el caso que demostramos) y $A' = UDV'$. Por lo tanto $A = VDU'$ (¡y la U' es la V' y la V es la U !). Por ejemplo, sea

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \text{ Acá } n \geq p, \text{ entonces } A'A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \text{ tiene como valores propios } 3 \text{ y } 1$$

asociados respectivamente a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ y $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Entonces

$$A = \begin{pmatrix} 2/3\sqrt{3} & 0 \\ 1/3\sqrt{3} & -1/\sqrt{2} \\ 1/3\sqrt{3} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Plan

- 1 Repaso de Probabilidad
- 2 Estimación
- 3 Tests de hipótesis
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos**
 - Estudio descriptivo
 - Matriz de varianzas y covarianzas
 - Correlación entre variables
 - Eliminación de variables redundantes lineales
- 7 Comparación de variables

- Las estadísticas están en todos lados: sondeos, noticias, medicina, biología, etc. Cuando se está en presencia de datos se puede hacer estadística.
 - 1 **Estadística descriptiva:** describir, representar, estructurar, resumir, sintetizar los datos (análisis de datos - Data Mining)
 - 2 **Estadística inferencial:** entender los datos, modelar los datos a partir de una función matemática y tomar decisiones. El objetivo consiste principalmente en extender las propiedades observadas en una muestra a toda la población (Machine Learning).
- Población, muestra, observación
- Las variables son las características de un individuo y pueden ser
 - **Cuantitativas**
 - continua: edad, precio, concentración ozono
 - discreta: nombres de especie, cantidad de veces que va al cine por mes...
 - **Cualitativas**
 - ordinal: nivel de satisfacción, intensidad viento
 - nominal: 0/1, spam/non spam, hincha de qué cuadro

Cuidado que muchas de las estadísticas descriptivas que se verán a continuación tienen sentido o no según el tipo de variable.

Matriz de datos

Hay dos maneras de ver la matriz de datos.

$\mathbf{X} = ((x_{ij}))_{i=1, \dots, n}^{j=1, \dots, p} \in \mathcal{M}_{n \times p}$ matriz de datos (n datos con p variables).

Por filas (individuos):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \text{individuo 1} \\ \text{individuo 2} \\ \vdots \\ \text{individuo n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_n \end{pmatrix}$$

Por columnas (características):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} v & v & & v \\ a & a & & a \\ r & r & & r \\ i & i & \dots & i \\ a & a & & a \\ b & b & & b \\ l & l & & l \\ e & e & & e \\ 1 & 2 & & p \end{pmatrix} = (x_1 \quad x_2 \quad \dots \quad x_p)$$

Matriz de datos

Datos Iris (salida de R):

```

Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1           3.5           1.4           0.2 setosa
2           4.9           3.0           1.4           0.2 setosa
3           4.7           3.2           1.3           0.2 setosa
4           4.6           3.1           1.5           0.2 setosa
5           5.0           3.6           1.4           0.2 setosa
6           5.4           3.9           1.7           0.4 setosa

```

Datos Airquality (salida de R):

```

Ozone Solar.R Wind Temp Month Day
1    41    190  7.4  67     5   1
2    36    118  8.0  72     5   2
3    12    149 12.6  74     5   3
4    18    313 11.5  62     5   4
5    NA     NA 14.3  56     5   5
6    28     NA 14.9  66     5   6

```

Matriz de datos

Para cada una de las variables (columnas) $j = 1, \dots, p$ tenemos estadísticos de descripción:

- Media de la variable j : $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \in \mathbb{R}$
- Varianza de la variable j : $s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$
- Desviación estándar de la variable j : s_j
- El mínimo, el máximo, el rango $|x_{j,max} - x_{j,min}|$
- La moda

El vector de medias de X es el vector $\bar{\mathbf{x}}$ en el cual cada coordenada de $\bar{\mathbf{x}}$ es la media de la columna j , es decir:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} = \frac{1}{n} \mathbf{X}' \mathbf{1} \in \mathbb{R}^p$$

donde $\mathbf{1}$ es el vector de \mathbb{R}^n con todas sus entradas iguales a 1.

Observar que:

- $\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) = \mathbf{0}_{\mathbb{R}^p}$.
- si $\bar{x}_j = 0$ entonces $s_j = \sqrt{\frac{1}{n} \|x_j\|^2}$

Serie estadística continua

Sea $\widehat{F}(x)$ la proporción de individuos de la muestra con un valor inferior o igual a x (la distribución empírica en x).

- La *mediana* es el valor x_m tal que $\widehat{F}(x_m) = 0,5$
- *Cuantiles*: $\widehat{F}(Q_1) = 0,25$, $\widehat{F}(Q_2) = 0,5$, $\widehat{F}(Q_3) = 0,75$
- Intervalo intercuantiles: $|Q_3 - Q_1|$

Ejemplo

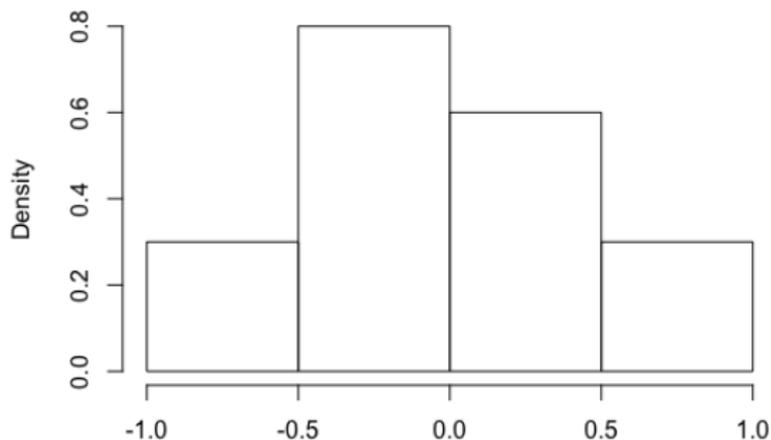
Datos:

```
0.07  0.45  0.78  0.85  0.49 -0.72 -0.36  0.55 -0.34  0.57
0.38  0.44 -0.92  0.49  0.84 -0.18 -0.85  0.31  0.00 -0.57
```

Datos ordenados:

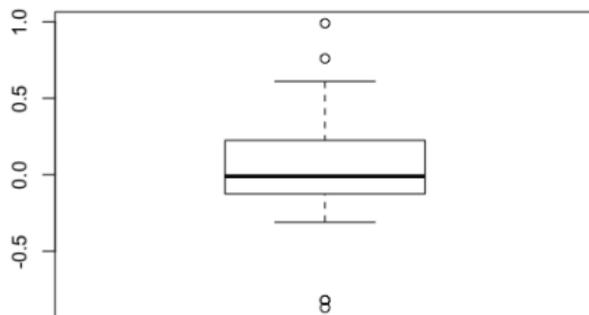
```
-0.87 -0.82 -0.82 -0.31 -0.13 -0.12 -0.11 -0.08 -0.03 -0.01
-0.01  0.03  0.08  0.21  0.22  0.23  0.40  0.61  0.76  0.99
```

Histograma:



Ejemplo

Boxplot:



- El valor indicado por la línea negra en el medio de la caja corresponde a la mediana de los datos
- Las extremidades de la caja a los cuartiles Q_1 y Q_3 .
- Los barras horizontales alrededor de la caja dan el mayor valor inferior a $Q_3 + 1,5(Q_3 - Q_1)$ y el menor valor superior a $Q_1 - 1,5(Q_3 - Q_1)$.
- Los puntos señalan valores extremos de los datos (posibles outliers) fuera de este intervalo $[Q_1 - 1,5(Q_3 - Q_1), Q_3 + 1,5(Q_3 - Q_1)]$

Parámetros de posición y de dispersión

$$X = \begin{pmatrix} 1 & 3,5 \\ 2 & 3,5 \\ 3 & 3,5 \\ 4 & 3,5 \\ 5 & 3,5 \\ 6 & 3,5 \end{pmatrix}$$

1 Parámetros de posición

- Media o promedio: $\bar{x} = (3,5, 3,5)'$
- Mediana: $\bar{\mathbf{m}}' = \left(\frac{3+4}{2}, \frac{3,5+3,5}{2} \right) = (3,5, 3,5)'$

2 Parámetros de dispersión

- Dispersión: $d' = \left(x_1^{(max)} - x_1^{(min)}, x_2^{(max)} - x_2^{(min)} \right)' = (6 - 1, 3,5 - 3,5)' = (5, 0)'$
- Varianza: $\mathbf{v} = \left(\frac{1}{6} \sum_{i=1}^6 (x_{i1} - \bar{x}_1)^2, \frac{1}{6} \sum_{i=1}^6 (x_{i2} - \bar{x}_2)^2 \right)' = (2,916, 0)'$
- Desvío: es la raíz cuadrada de la varianza.

Centrar y reducir la matriz de datos

Decimos que los datos de una variable son *centrados* si les sacamos su media y son *centrados/reducidos* si además los dividimos por su desvío. Los datos de una variable centrada y reducida son útiles porque no tienen más unidades y datos provenientes de variables distintas pueden ser comparables.

$$x_{ij} \longrightarrow z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

$$X = \begin{pmatrix} 1 & 3,5 \\ 2 & 3,5 \\ 3 & 3,5 \\ 4 & 3,5 \\ 5 & 3,5 \\ 6 & 3,5 \end{pmatrix} \longrightarrow Z = \begin{pmatrix} -1,464 & 0 \\ -0,878 & 0 \\ -0,293 & 0 \\ 0,293 & 0 \\ 0,878 & 0 \\ 1,464 & 0 \end{pmatrix}$$

Cuidado que esto tiene sentido si todas las variables son continuas.

Matriz de datos centrados

La misma se obtiene de la matriz de datos original restando en cada columna la media de la

variable correspondiente. Formalmente, si $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{M}_{n \times 1}$, la matriz de datos centrados es:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}' = \mathbf{X} - \frac{1}{n}\mathbf{1}(\mathbf{1}'\mathbf{X}) = \underbrace{\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}' \right)}_{=P} \mathbf{X} = P\mathbf{X}$$

Entonces P es simétrica, idempotente ($PP = P$) y de rango $= n - 1$ (pues es ortogonal al subespacio generado por el vector $\mathbf{1}$).

Además:

$$S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \frac{1}{n}\mathbf{X}'P\mathbf{X}$$

Observación: Se puede definir la matriz de varianzas corregida $\hat{S} = \frac{1}{n-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ ($n - 1$ en lugar de n) a los efectos de tener un estimador insesgado de la matriz de la población. Este cambio en la división tiene el mismo origen que en el caso univariado: hay $n - 1$, y no n , desviaciones estandares independientes ya que los vectores de desviación están ligados por $\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) = 0$.

Matriz de varianzas y covarianzas

La relación **lineal** entre dos variables se mide por la covarianza. La covarianza entre las variables j y k se define como:

$$\text{Cov}(x_j, x_k) = s_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{1}{n} \left\langle \begin{pmatrix} x_{1j} - \bar{x}_j \\ x_{2j} - \bar{x}_j \\ \vdots \\ x_{nj} - \bar{x}_j \end{pmatrix}, \begin{pmatrix} x_{1k} - \bar{x}_k \\ x_{2k} - \bar{x}_k \\ \vdots \\ x_{nk} - \bar{x}_k \end{pmatrix} \right\rangle$$

- Si suponemos que $\bar{x}_j = \bar{x}_k = 0$ y que $s_j = s_k = 1$ (matriz centrada y reducida) entonces:

$$x'_j x_k = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}} = r_{jk}$$

donde r_{jk} es el coeficiente de correlación de las variables j y k .

- Si suponemos que $\bar{x}_j = \bar{x}_k = 0$ se tiene que:

$$\cos(x_j, x_k) = \frac{x'_j x_k}{\|x'_j\| \|x_k\|} = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}} = r_{jk}$$

Matriz de varianzas y covarianzas

Propiedades de la covarianza

① $cov(x_j, x_k) = \frac{1}{n} \langle x_j, x_k \rangle - \bar{x}_j \bar{x}_k$

② $cov(x_j, x_k)$ es el coeficiente (j, k) de la matriz de varianzas covarianzas

$$S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

③ $cov(x_j, x_k) = cov(x_k, x_j)$.

④ $cov(x_j, x_j) = var(x_j) = s_j^2$

Matriz de varianzas y covarianzas

Propiedades de la covarianza

- ① $cov(x_j, x_k) = \frac{1}{n} \langle x_j, x_k \rangle - \bar{x}_j \bar{x}_k$
 - ② $cov(x_j, x_k)$ es el coeficiente (j, k) de la matriz de varianzas covarianzas
- $$S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$
- ③ $cov(x_j, x_k) = cov(x_k, x_j)$.
 - ④ $cov(x_j, x_j) = var(x_j) = s_j^2$

Propiedades de la correlación

- ① $r_{jj} = 1$ para todo $j = 1, \dots, p$.
- ② $-1 \leq r_{jk} \leq 1$ para todo $j, k = 1, \dots, p$.
- ③ $|r_{jk}| = 1$ si y solo si existe $\lambda \in \mathbb{R}$ tal que $x_j - \bar{x}_j = \lambda(x_k - \bar{x}_k)$

Importante

- Si la correlación está cerca de 1, la misma implica una relación lineal entre los datos pero no necesariamente una causa. Por ejemplo en invierno la cantidad de resfriados de una población puede ser correlada con la cantidad de garrafas vendidas pero ningún hecho es causante del otro.
- Si la correlación está cerca de 0, esto no significa que no haya relación entre las variables, solo que no es lineal. Puede ser de otro tipo (cuadrática, logarítmica, etc.)

Matriz de varianzas y covarianzas

Consideramos distintas matrices:

- **Matriz de datos:**

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_n \end{pmatrix} \in \mathcal{M}_{n \times p}$$

- **Matriz de varianzas y covarianzas:**

$$S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix} \in \mathcal{M}_{p \times p}$$

- **Matriz de correlaciones:**

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix} \in \mathcal{M}_{p \times p}$$

siendo $r_{ij} = \frac{s_{ij}}{s_i s_j}$.

Matriz de varianzas y covarianzas en R

Un ejemplo en R. Datos de 150 flores de lirio.

- Adjunto los datos: `attach(iris)`
- Considero la parte de la base con las variables independientes `x=iris[,1:4]`
- Cálculo la media de cada variable: `colMeans(x)`
- Cálculo la varianza o el desvío de cada variable: `apply(x,2,var)` ; `apply(x,2,sd)`
- La matriz de varianzas y covarianzas es: `S=cov(x)` o `S=var(x)`
- La matriz de correlaciones es: `R=cor(x)`
- Buscar los valores y vectores propios de `S`: `eigen(S)`

Matriz de varianzas y covarianzas - Propiedades

Veamos una serie de propiedades de la matriz de varianzas y covarianzas

$$S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix} \in \mathcal{M}_{p \times p}$$

- 1 S es simétrica y tiene el mismo rango que $\tilde{\mathbf{X}}$.
- 2 S es diagonalizable en una base ortonormal de \mathbb{R}^p .
- 3 S es semidefinida positiva, o sea para todo $w \in \mathbb{R}^p$ $w' S w \geq 0$.
- 4 S tiene todos sus valores propios ≥ 0 .
- 5 $\det(S) \geq 0$, $tr(S) \geq 0$.

Eliminación de variables redundantes lineales

Veamos como es posible reducir la cantidad de variables de la matriz de datos cuando una variable es combinación lineal de otras.

Lema: $\{w \in \mathbb{R}^p : w' S w = 0\} = N(S)$.

Proposición: $N(S) = N(\tilde{\mathbf{X}})$

Dem:

$$\begin{aligned} w \in N(S) &\Leftrightarrow S w = 0_{\mathbb{R}^p} \Leftrightarrow w' \tilde{\mathbf{X}}' \tilde{\mathbf{X}} w = 0 \\ &\Leftrightarrow (\tilde{\mathbf{X}} w)' \tilde{\mathbf{X}} w = 0 \Leftrightarrow \|\tilde{\mathbf{X}} w\|^2 = 0 \Leftrightarrow \tilde{\mathbf{X}} w = 0_{\mathbb{R}^n} \Leftrightarrow w \in N(\tilde{\mathbf{X}}) \end{aligned}$$

Conclusión: si un vector w se encuentra en el núcleo de S , sus coordenadas son los coeficientes de la combinación lineal existente entre las variables de la matriz de datos.

Eliminación de variables redundantes: ejemplo

Consideremos la siguiente matriz de varianzas y covarianzas (ejemplo 3.3 del libro de Peña):

$$S = \begin{pmatrix} 0,0947 & 0,0242 & 0,0054 & 0,0594 \\ 0,0242 & 0,0740 & 0,0285 & 0,0491 \\ 0,0054 & 0,0285 & 0,0838 & 0,0170 \\ 0,0594 & 0,0285 & 0,0170 & 0,0543 \end{pmatrix}$$

cuyos valores propios son 0.172, 0.08762, 0.0461 y 0.00005 (el rango es aprox. 3).

Un vector propio asociado al valor propio cero es $(0,408, 0,408, 0, -0,816)$ que se puede reescribir como $(0,5, 0,5, 0, -1)$ por lo que la 4ta variable de X es el promedio de las dos primeras.

Obs.: El razonamiento anterior se puede extender para cualquier número de valores propios nulos: si S tiene rango h entonces existen $p - h$ variables redundantes que podemos eliminar y los vectores asociados a estos valores propios indican la CL de estas variables redundantes (ver páginas 76 y 77 del libro de Peña).

Plan

- 1 Repaso de Probabilidad
- 2 Estimación
- 3 Tests de hipótesis
- 4 Tests de normalidad
- 5 Repaso de Algebra Lineal
- 6 Matriz de datos
- 7 Comparación de variables**
 - Relación entre dos series cuantitativas
 - Relación entre dos series cualitativas

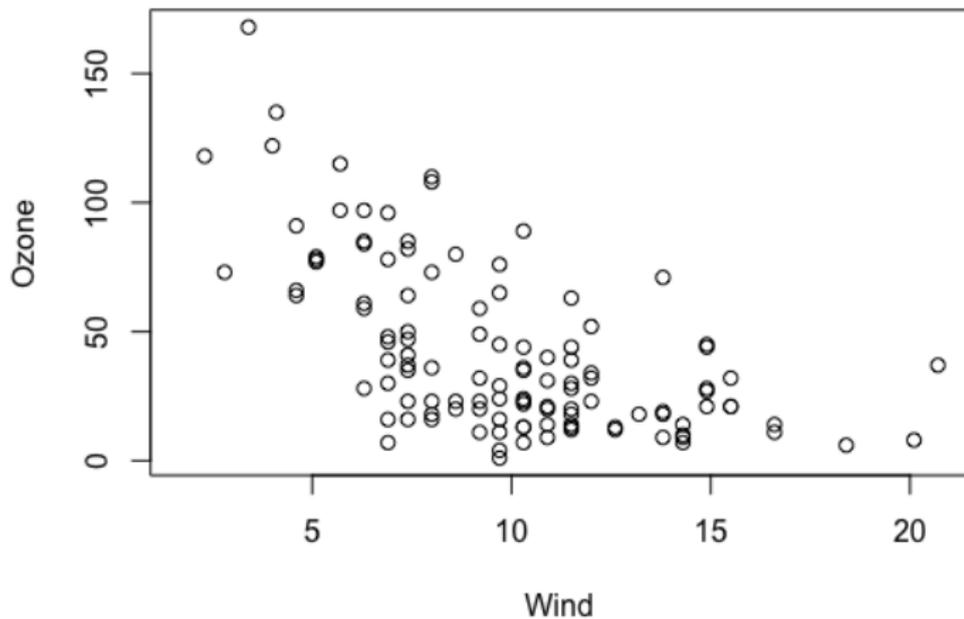
Relación entre dos series estadísticas

Se observa sobre n individuos dos caracteres (x_1, \dots, x_n) e (y_1, \dots, y_n) . correspondientes a dos variables x e y . Hay tres casos:

- Las dos variables son cuantitativas (continuas)
- Las dos variables son cualitativas (discretas)
- Una es cuantitativa y la otra es cualitativa

Dos series continuas

Dibujar la nube de puntos

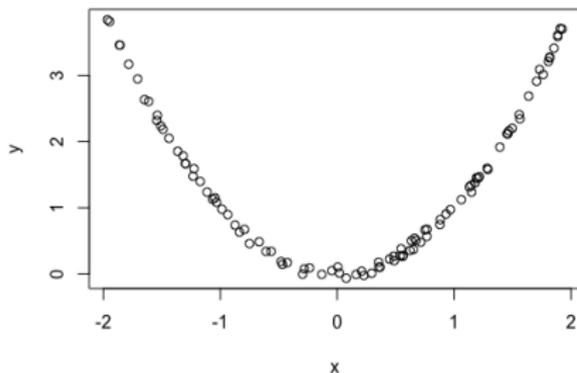


Dos series continuas

Correlación lineal:

$$\rho_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{s_x s_y}$$

cor= 0.069



Si la correlación está cerca de 0, esto no significa que no haya relación entre las variables, solo que no es lineal. Puede ser de otro tipo (cuadrática, logarítmica, etc).

Dos series cualitativas

Podemos querer conocer la opinión de consumidores de tv cable en un barrio de Montevideo en función del cable que tienen, I posibilidades, y su conformidad en cuanto a la programación, J posibles opiniones, obteniéndose de esta manera una matriz $X \in \mathcal{M}_{I \times J}$.

| Cable | Poco Conforme | Conforme | Muy conforme | Total |
|-------------|---------------|----------|--------------|-------|
| Nuevo Siglo | 3 | 47 | 178 | 228 |
| Montecable | 24 | 56 | 20 | 100 |
| TCC | 12 | 8 | 23 | 43 |
| DirectTV | 2 | 14 | 88 | 104 |
| Total | 41 | 125 | 309 | 475 |

$$X = \begin{pmatrix} 3 & 47 & 178 \\ 24 & 56 & 20 \\ 12 & 8 & 23 \\ 2 & 14 & 88 \end{pmatrix} \in \mathcal{M}_{4 \times 3}$$

Cada persona aparece en una sola casilla de la tabla.

Tabla de contingencias

| $X Y$ | y_1 | \dots | y_j | \dots | y_J | |
|-------|---------|---------|----------|---------|---------|----------|
| x_1 | | | \vdots | | | |
| x_i | \dots | \dots | n_{ij} | \dots | \dots | $n_{i.}$ |
| x_I | | | \vdots | | | |
| | | | $n_{.j}$ | | | |

Cuadro: Tabla de contingencias

- Suponemos que $I > J$.
- x_1, x_2, \dots, x_I representan las modalidades de la variables X .
- y_1, y_2, \dots, y_J representan las modalidades de la variables Y .
- n es la cantidad de individuos
- n_{ij} es la cantidad de individuos que cumplen la modalidad i de la variable X y la modalidad j de la variable Y .
- $n_{i.}$ es la cantidad de individuos que cumplen la modalidad i de la variable X
- $n_{.j}$ es la cantidad de individuos que cumplen la modalidad j de la variable Y .

ejemplo

| | Hotel | Locación | Res.Secund | Padres | Amigos | Camping | Grupo Viaje | Otros | Total |
|----------------|-------|----------|------------|--------|--------|---------|-------------|-------|-------|
| Prod. Rurales | 195 | 62 | 1 | 499 | 44 | 141 | 49 | 65 | 1056 |
| Jefes | 700 | 354 | 229 | 959 | 185 | 292 | 119 | 140 | 2978 |
| Ejecutivo sup | 961 | 471 | 633 | 1580 | 305 | 360 | 162 | 148 | 4620 |
| Ejecutivo prom | 572 | 537 | 279 | 1689 | 206 | 748 | 155 | 112 | 4298 |
| Empleado | 441 | 404 | 166 | 1079 | 178 | 434 | 178 | 92 | 2972 |
| Obrero | 783 | 1114 | 387 | 4052 | 497 | 1464 | 525 | 387 | 9209 |
| Otras prof. | 142 | 103 | 210 | 1133 | 132 | 181 | 46 | 59 | 2006 |
| Inactivos | 741 | 332 | 327 | 1789 | 311 | 236 | 102 | 102 | 3940 |
| Total | 4535 | 3377 | 2232 | 12780 | 1858 | 3856 | 1336 | 1105 | 31079 |
| | | | | | | | | | |
| | | | | | | | | | |

Matriz de frecuencias relativas

Podemos trabajar con la matriz F de frecuencias relativas, pensada como matriz de probabilidades, que se obtiene de la matriz anterior dividiendo cada casilla de la tabla de contingencia por n el total de valores observados

| $X Y$ | y_1 | \dots | y_j | \dots | y_J | |
|-------|---------|---------|---------------------------------------|---------|---------|-------------------------------------|
| x_1 | | | \vdots | | | |
| x_i | \dots | \dots | $f_{ij} = \frac{n_{ij}}{n}$ | \dots | \dots | $f_{i\cdot} = \frac{n_{i\cdot}}{n}$ |
| x_I | | | \vdots | | | |
| | | | $f_{\cdot j} = \frac{n_{\cdot j}}{n}$ | | | |

- A $f_{i\cdot}$ se le llama *frecuencia marginal de la modalidad i* .
- A $f_{\cdot j}$ se le llama *frecuencia marginal de la modalidad j* .
- A $f_i^j = \frac{n_{ij}}{n_{i\cdot}} = \frac{f_{ij}}{f_{i\cdot}}$ se le llama *frecuencia condicional j sabiendo i* .
- A $f_j^i = \frac{n_{ij}}{n_{\cdot j}} = \frac{f_{ij}}{f_{\cdot j}}$ se le llama *frecuencia condicional i sabiendo j* .

Distancia entre filas

Cada fila puede considerarse como un punto en el espacio euclideo \mathbb{R}^J . La idea consistirá en buscar un subespacio de \mathbb{R}^J de dimensión menor donde podamos ver la distancia entre estos puntos, de manera que las filas que tienen estructura parecidas estén cercas y las que tienen estructuras muy distintas alejadas.

La distancia euclidea no es una buena medida de proximidad de los puntos en la matriz de frecuencia F :

| | | | | | |
|---|------|------|------|------|-----|
| A | 0.03 | 0.06 | 0.15 | 0.06 | 0.3 |
| B | 0.07 | 0.14 | 0.35 | 0.14 | 0.7 |
| T | 0.1 | 0.2 | 0.5 | 0.2 | 1 |

Si hacemos la distancia euclidea entre las filas obtenemos un valor alto. Sin embargo las dos filas tienen exactamente la misma estructura relativa pues si dividimos cada casillero por la frecuencia relativa de la fila f_i . obtenemos:

| | | | | | |
|---|-----|-----|-----|-----|---|
| A | 0.1 | 0.2 | 0.5 | 0.2 | 1 |
| B | 0.1 | 0.2 | 0.5 | 0.2 | 1 |

La operación matricial para pasar de una tabla a otra es

$$R = D_f^{-1}F$$

donde D_f es una matriz diagonal $I \times I$ que tiene en su diagonal principal al vector

$$\mathbf{f} = (f_{1\cdot}, f_{2\cdot}, \dots, f_{I\cdot})$$

Esto motiva el poder considerar las tablas de los perfiles fila y columna.

Tabla de perfiles por filas

Podemos trabajar con la matriz de perfiles filas que se obtiene de la tabla de contingencia dividiendo cada casilla de la fila i por la suma n_i . del total de valores observados para ella.

| $X Y$ | y_1 | y_2 | \dots | y_j | \dots | y_J | |
|---------|-------------------------|-------------------------|---------|-------------------------|---------|-------------------------|----------|
| x_1 | $\frac{f_{11}}{f_{1.}}$ | $\frac{f_{12}}{f_{1.}}$ | \dots | $\frac{f_{1j}}{f_{1.}}$ | \dots | $\frac{f_{1J}}{f_{1.}}$ | 1 |
| \dots | \dots | \dots | \dots | \dots | \dots | \dots | \vdots |
| i | $\frac{f_{i1}}{f_{i.}}$ | $\frac{f_{i2}}{f_{i.}}$ | \dots | $\frac{f_{ij}}{f_{i.}}$ | \dots | $\frac{f_{iJ}}{f_{i.}}$ | 1 |
| \dots | \dots | \dots | \dots | \dots | \dots | \dots | \vdots |
| x_I | $\frac{f_{I1}}{f_{I.}}$ | $\frac{f_{I2}}{f_{I.}}$ | \dots | $\frac{f_{Ij}}{f_{I.}}$ | \dots | $\frac{f_{IJ}}{f_{I.}}$ | 1 |

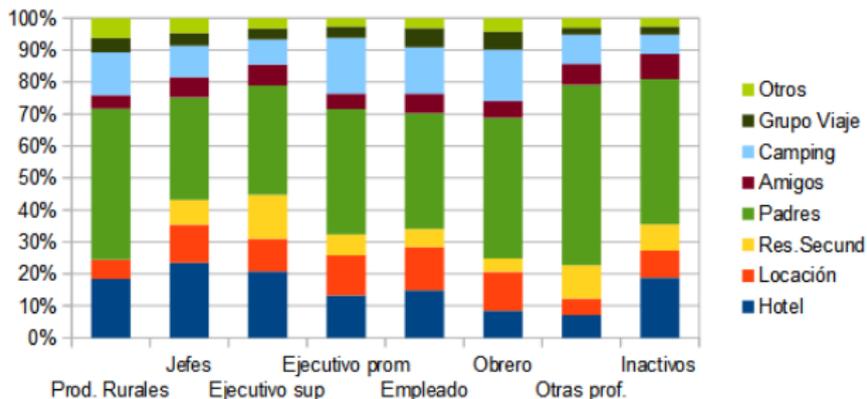
Cuadro: Tabla de perfiles filas; cada fila suma 1

La misma permite comparar la repartición de los valores de Y en las distintas modalidades de X .

Como todas las filas suman 1, todos los puntos i están sobre un hiperplano de dimensión $J - 1$ de \mathbb{R}^J .

Ejemplo

| | Hotel | Locación | Res.Second | Padres | Amigos | Camping | Grupo Viaje | Otros | Total |
|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| Prod. Rurale | 0,184659091 | 0,058712121 | 0,00094697 | 0,472537879 | 0,041666667 | 0,133522727 | 0,046401515 | 0,06155303 | 1 |
| Jefes | 0,235057085 | 0,118871726 | 0,076897246 | 0,322028207 | 0,06212223 | 0,098052384 | 0,039959704 | 0,047011417 | 1 |
| Ejecutivo sup | 0,208008658 | 0,101948052 | 0,137012987 | 0,341991342 | 0,066017316 | 0,077922078 | 0,035064935 | 0,032034632 | 1 |
| Ejecutivo pro | 0,133085156 | 0,124941833 | 0,064913913 | 0,392973476 | 0,047929269 | 0,174034435 | 0,036063285 | 0,026058632 | 1 |
| Empleado | 0,148384926 | 0,135935397 | 0,055854643 | 0,363055182 | 0,059892328 | 0,14602961 | 0,059892328 | 0,030955585 | 1 |
| Obrero | 0,085025519 | 0,120968618 | 0,042024107 | 0,440004344 | 0,053968943 | 0,158974916 | 0,057009447 | 0,042024107 | 1 |
| Otras prof. | 0,070787637 | 0,051345962 | 0,104685942 | 0,564805583 | 0,065802592 | 0,090229312 | 0,022931206 | 0,029411765 | 1 |
| Inactivos | 0,188071066 | 0,084263959 | 0,082994924 | 0,454060914 | 0,07893401 | 0,059898477 | 0,025888325 | 0,025888325 | 1 |
| Total | 1,253079138 | 0,796987669 | 0,565330733 | 3,351456926 | 0,476333356 | 0,938663939 | 0,323210747 | 0,294937493 | 8 |



El 23 % de los jefes van al hotel y en las otras profesiones el 56 % van a la casa de los padres.

Tabla de perfiles por columnas

Podemos trabajar con la matriz de perfiles columnas que se obtiene de la tabla de contingencia dividiendo cada casilla de la columna j por la suma $n_{.j}$ del total de valores observados para ella.

| $X Y$ | y_1 | \dots | y_j | \dots | y_J |
|-------|-------------------------|----------|-------------------------|----------|-------------------------|
| 1 | $\frac{f_{11}}{f_{.1}}$ | \vdots | $\frac{f_{1j}}{f_{.j}}$ | \vdots | $\frac{f_{1J}}{f_{.J}}$ |
| 2 | $\frac{f_{21}}{f_{.2}}$ | \vdots | $\frac{f_{2j}}{f_{.j}}$ | \vdots | $\frac{f_{2J}}{f_{.J}}$ |
| | \vdots | \vdots | \vdots | \vdots | \vdots |
| i | $\frac{f_{i1}}{f_{.1}}$ | \vdots | $\frac{f_{ij}}{f_{.j}}$ | \vdots | $\frac{f_{iJ}}{f_{.J}}$ |
| | \vdots | \vdots | \vdots | \vdots | \vdots |
| I | $\frac{f_{I1}}{f_{.1}}$ | \vdots | $\frac{f_{Ij}}{f_{.j}}$ | \vdots | $\frac{f_{IJ}}{f_{.J}}$ |
| | 1 | \dots | 1 | \dots | 1 |

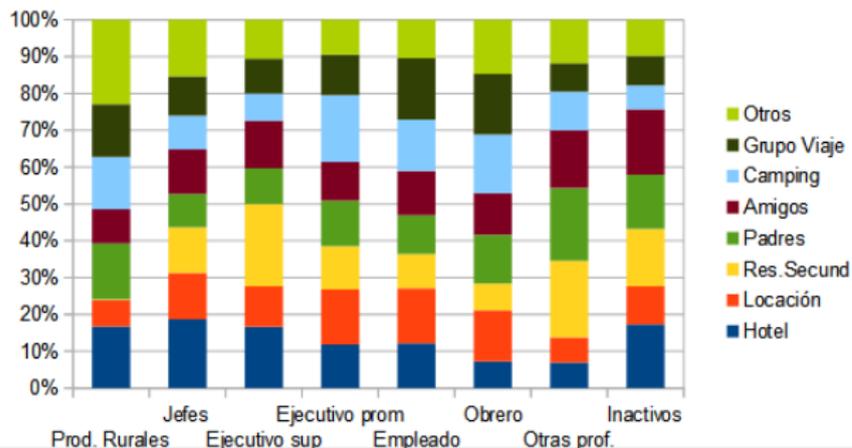
Cuadro: Tabla de perfiles columnas; cada columna suma 1.

La misma permite comparar la repartición de los valores de X en las distintas modalidades de Y .

Idem acá: todas las columnas suman 1, todos los puntos j están sobre un hiperplano de dimensión $I - 1$ de \mathbb{R}^I .

Ejemplo

| | Hotel | Locación | Res.Secund | Padres | Amigos | Camping | Grupo Viaje | Otros | Total |
|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Prod. Rurale | 0,042998897 | 0,018359491 | 0,000448029 | 0,039045383 | 0,023681378 | 0,03656639 | 0,036676647 | 0,058823529 | 0,256599744 |
| Jefes | 0,154355017 | 0,104826769 | 0,102598566 | 0,075039124 | 0,099569429 | 0,075726141 | 0,089071856 | 0,126696833 | 0,827883735 |
| Ejecutivo sup | 0,211907387 | 0,139472905 | 0,283602151 | 0,123630673 | 0,164155005 | 0,093360996 | 0,121257485 | 0,133936652 | 1,271323253 |
| Ejecutivo pro | 0,126130099 | 0,159016879 | 0,125 | 0,132159624 | 0,110871905 | 0,193983402 | 0,116017964 | 0,101357466 | 1,06453734 |
| Empleado | 0,09724366 | 0,11963281 | 0,07437276 | 0,084428795 | 0,095801938 | 0,112551867 | 0,133233533 | 0,083257919 | 0,800523282 |
| Obrero | 0,172657111 | 0,32987859 | 0,173387097 | 0,317057903 | 0,267491927 | 0,37966805 | 0,392964072 | 0,350226244 | 2,383330994 |
| Otras prof. | 0,031312018 | 0,030500444 | 0,094086022 | 0,088654147 | 0,071044133 | 0,046939834 | 0,034431138 | 0,053393665 | 0,450361401 |
| Inactivos | 0,16339581 | 0,098312111 | 0,146505376 | 0,139984351 | 0,167384284 | 0,06120332 | 0,076347305 | 0,092307692 | 0,94544025 |
| Total | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |



El 15,4 % de las personas que van al hotel son jefes. Dentro de las personas que van al hotel hay una mayoría de ejecutivos superiores, pero estos últimos prefieren ir en lo de los padres (ver perfil filas)

Prueba de independencia entre las variables

Antes de comenzar el estudio debemos ver si las dos variables son independientes.

Recordamos que dos variables X e Y son independientes si para todo par (i, j) se tiene que

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) (*)$$

Esto equivale, cuando se puede hablar de probabilidad condicional, a que

- para todo par (i, j) , $\mathbb{P}(X = x_i | Y = y_j) = \mathbb{P}(X = x_i)$.
- para todo par (i, j) , $\mathbb{P}(Y = y_j | X = x_i) = \mathbb{P}(Y = y_j)$.

La expresión (*) que se traduce en la tabla de frecuencias relativas por

$$f_{ij} = f_i \cdot f_j \quad \forall i = 1, \dots, I, j = 1, \dots, J.$$

Si no se cumple esta igualdad para todo i y para todo j hay algún grado de asociación entre ambas variables.

Observar que

$$\frac{f_{ij}}{f_i} = \frac{n_{ij}}{n_i}$$

por lo que la propiedad de independencia se traduce como

$$\frac{n_{ij}}{n} = \frac{n_i}{n} \frac{n_j}{n} \quad \forall i, j \Rightarrow \underbrace{n_{ij}}_{\text{valor observado}} = \underbrace{\frac{n_i \cdot n_j}{n}}_{\text{valor teórico}} \quad \forall i, j$$

Por último:

- Si $f_{ij} > f_i \cdot f_j$ decimos que las modalidades i y j se atraen.
- Si $f_{ij} < f_i \cdot f_j$ decimos que las modalidades i y j se repelen.

Prueba de independencia entre las variables

Se hace necesario definir un test estadístico global que mida de alguna manera la distancia entre lo observado y lo que uno espera, dado que se cumple la hipótesis nula de independencia entre las variables.

- (H0): X e Y son independientes
 (H1): X e Y no son independientes

El estadístico es:

$$\begin{aligned}\chi^2 &= \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n}\right)^2}{\frac{n_{i.} \cdot n_{.j}}{n}} = \sum_{i=1}^I \sum_{j=1}^J \frac{(nf_{ij} - nf_{i.} \cdot f_{.j})^2}{nf_{i.} \cdot f_{.j}} \\ &= n \sum_{i=1}^I \sum_{j=1}^J \frac{(\text{fr. observadas} - \text{fr esperadas})^2}{\text{fr. esperadas}}\end{aligned}$$

El valor n_{ij} es el valor observado en la celda i/j y el cociente $\frac{n_{i.} \cdot n_{.j}}{n}$ es el valor esperado de la celda ij bajo (H0). El estadístico χ^2 mide el desvío entre lo que se observa y lo esperado en caso de independencia y sigue asintóticamente una distribución χ^2 con $(I - 1) \times (J - 1)$ grados de libertad.

Si el valor de χ^2 es grande entonces tenemos una dependencia grande entre las variables. Sin embargo al ser un indicador global, es insuficiente para medir las asociaciones entre las modalidades.

Prueba de independencia entre las variables

Veamos en el ejemplo siguiente por qué se divide por el valor teórico.

$$\frac{\left(n_{ij} - \frac{n_i \cdot n_j}{n}\right)^2}{\frac{n_i \cdot n_j}{n}}$$

Jugamos a cara o cruz 10 veces. Se gana 1 vez y la diferencia es 4. Jugamos a cara o cruz 100 veces. Se gana 46 veces y la diferencia es 4.

| Teórico | Observado | Diferencia | Diferencia ² | $\frac{\text{Diferencia}^2}{\text{Teórico}}$ |
|---------|-----------|------------|-------------------------|--|
| 5 | 1 | 4 | 16 | 3.2 |
| 50 | 46 | 4 | 16 | 0.32 |

Claramente no es lo mismo ganar 1 vez en 10 que ganar 46 veces en 100.

Ejemplo

```
> base=read.table("vacaciones.csv",sep=",",header=TRUE,row.names=1)
> k=chisq.test(base)
> k$observed
```

| | Hotel | Locación | Res.Secund | Padres | Amigos | Camping | Grupo.Viaje | Otros |
|----------------|-------|----------|------------|--------|--------|---------|-------------|-------|
| Prod. Rurales | 195 | 62 | 1 | 499 | 44 | 141 | 49 | 65 |
| Jefes | 700 | 354 | 229 | 959 | 185 | 292 | 119 | 140 |
| Ejecutivo sup | 961 | 471 | 633 | 1580 | 305 | 360 | 162 | 148 |
| Ejecutivo prom | 572 | 537 | 279 | 1689 | 206 | 748 | 155 | 112 |
| Empleado | 441 | 404 | 166 | 1079 | 178 | 434 | 178 | 92 |
| Obrero | 783 | 1114 | 387 | 4052 | 497 | 1464 | 525 | 387 |
| Otras prof. | 142 | 103 | 210 | 1133 | 132 | 181 | 46 | 59 |
| Inactivos | 741 | 332 | 327 | 1789 | 311 | 236 | 102 | 102 |

```
> k$expected
```

| | Hotel | Locación | Res.Secund | Padres | Amigos | Camping | Grupo.Viaje | Otros |
|----------------|-----------|-----------|------------|-----------|-----------|-----------|-------------|-----------|
| Prod. Rurales | 154.0899 | 114.7435 | 75.83873 | 434.2379 | 63.13099 | 131.0189 | 45.39451 | 37.54561 |
| Jefes | 434.5452 | 323.5853 | 213.87097 | 1224.5838 | 178.03417 | 369.4832 | 128.01596 | 105.88146 |
| Ejecutivo sup | 674.1433 | 502.0026 | 331.79446 | 1899.7909 | 276.19808 | 573.2076 | 198.60098 | 164.26204 |
| Ejecutivo prom | 627.1576 | 467.0146 | 308.66939 | 1767.3812 | 256.94791 | 533.2568 | 184.75910 | 152.81348 |
| Empleado | 433.6697 | 322.9333 | 213.44007 | 1222.1165 | 177.67547 | 368.7388 | 127.75804 | 105.66814 |
| Obrero | 1343.7632 | 1000.6369 | 661.36259 | 3786.8342 | 550.54287 | 1142.5691 | 395.86937 | 327.42189 |
| Otras prof. | 292.7124 | 217.9691 | 144.06487 | 824.8875 | 119.92497 | 248.8863 | 86.23238 | 71.32244 |
| Inactivos | 574.9188 | 428.1148 | 282.95891 | 1620.1680 | 235.54555 | 488.8394 | 169.36967 | 140.08494 |

```
> k
```

Pearson's Chi-squared test

```
data: base
```

```
X-squared = 2292.148, df = 49, p-value < 2.2e-16
```

En la primera tabla vemos los n_{ij} y en la segunda tabla los productos $n_{i.}n_{.j}/n$.

Conclusión: Hay una dependencia significativa entre las variables.