FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

EYP1113 - Probabilidad y Estadística

Taller R - 07: Distribución conjunta: Caso continuo

Ricardo Aravena C. - Cristian Capetillo C. - Ingrid Guevara R.
Ricardo Olea O. - Bladimir Morales T., Daniel Saavedra M.

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2024

Contenido I

Motivación

Distribución Normal bivariada

Aplicación

Motivación

Con la excepción del taller 06, nuestro enfoque a estado puesto en el análisis univariado, es decir, el análisis de cada variable por separado. No obstante, y cómo se vió en la clase anterior, un análisis puede conllevar a una mejor comprensión de la realidad cuando se trabajan las variables conjuntamente.

Imaginemos estamos interesados en estudiar el peso (X_1) y la altura (X_2) de cierto grupo de personas. El enfoque conjunto nos abre un montón de interpretaciones nuevas, siendo posible calcular probabilidades tales como

$$\mathbb{P}(X_1 > 50kg | X_2 > 1.4mts).$$

En este taller nos enfocaremos en el análisis conjunto de variables continuas. En particular, aplicaremos la distribución Normal bivariada.

Motivación

Al igual que en el taller anterior, se usará la base de datos de `heightweight` disponible en el paquete `gcookbook`, que contiene información de ciertas características de 236 personas, como la altura, sexo, peso y otros.

Nuestro interés estará puesto solo en el peso y la altura, aunque podríamos considerar todas las variables que fuesen necesarias.

Echemos un vistazo a los datos.

Distribución conjunta, marginal y condicional

Sean X_1 y X_2 variables aleatorias continuas. Si definimos la función de densidad **conjunta** mediante $f_{X_1,X_2}(x_1,x_2)$ se tiene que:

- La función de densidad **marginal** de X_1 y X_2:

$$f_{X_1}(x_1) = \int f_{X_1,X_2}(x_1,x_2)dx_2 \quad ; \quad f_{X_2}(x_2) = \int f_{X_1,X_2}(x_1,x_2)dx_1.$$

- La función de densidad **condicional** de X_1 dado $X_2 = x_2$ e X_2 dado $X_1 = x_1$:

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} \quad ; \quad f_{X_2|X_1=x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}.$$

Distribución Normal bivariada

Dos variables, X_1 y X_2, distribuyen conjuntamente una Normal bivariada con vector de medias y matriz de covarianzas

$$\boldsymbol{\mu} = (\mu_1, \mu_2)^T \quad , \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$

denotado por $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$, si su función de densidad **conjunta** está dada por

$$f_{X_1,X_2}(x_1, x_2) = (2\pi)^{-1}|\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\},$$

donde $\boldsymbol{x} = (x_1, x_2)^T$.

Distribución Normal bivariada

La distribución **marginal** de X_1 es $\mathcal{N}(\mu_1, \sigma_1^2)$ y la de X_2, $\mathcal{N}(\mu_2, \sigma_2^2)$.

Por otro lado, la distribución **condicional** de X_1 dado $X_2 = x_2$ es una $\mathcal{N}(\mu_{1|2}, \sigma_{1|2}^2)$, donde

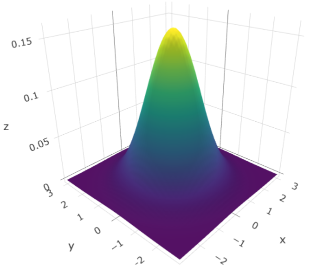
$$\mu_{1|2} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),$$
$$\sigma_{1|2}^2 = \sigma_1^2(1 - \rho^2),$$

con $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ correspondiente a la correlación entre X_1 y X_2.

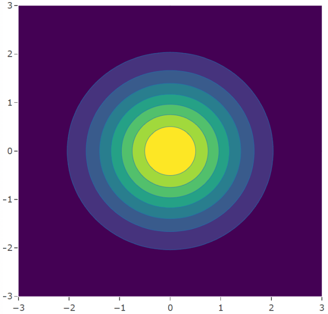
Análogamente, $X_2|X_1 = x_1 \sim \mathcal{N}(\mu_{2|1}, \sigma_{2|1}^2)$ con

$$\mu_{2|1} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),$$
$$\sigma_{2|1}^2 = \sigma_2^2(1 - \rho^2).$$

Gráficos



(a) Densidad $\mathcal{N}_2(\mathbf{0}, I)$.



(b) Curvas de nivel $\mathcal{N}_2(\mathbf{0}, I)$.

Aplicación

Con la base de datos `heightweight`, supongamos que el peso (X_1) y la altura (X_2) distribuyen normal bivariado, con vector de medias y matriz de covarianzas, en principio desconocidas, reemplazadas por sus análogos muestrales.

Comparemos lo calculado en la muestra, con lo que sería en la población (si nuestros supuestos son ciertos).