Muestreo y geometría

Monday, February 7, 2022 10:44 PM

Muestreo y geometría

Asumiremos muestras aleatorias

$X = \begin{bmatrix} x_{11} & x_{1p} \\ x_{n1} & x_{np} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_n' \end{bmatrix} \begin{matrix} \to \text{primera obs.} \\ \to \text{última obs.} \end{matrix}$

c/ fila es una obs.

n vecs en $\mathbb{R}^p$

→ c/ ejes una var.

se pueden ver como n puntos en $\mathbb{R}^p$ $(p \leq 3)$

$\bar{x}$ es el centro de gravedad de la "nube" de puntos.

Una representaº geom.

- P - vectores en el espacio n-dim.

$\underset{n \times p}{X} = \begin{bmatrix} x_{11} & x_{1p} \\ x_{n1} & x_{np} \end{bmatrix} = [y_1 \dots y_p]$

→ y_i es un vec. de dim n $i = 1 \dots p$

$y_i = \begin{bmatrix} x_{1i} \\ x \end{bmatrix}$ son las n mediciones de la var $i = 1, \dots, p$ — si $n > 3$, no se puede visualizar.

└ tiene todas las obs. de la var i.

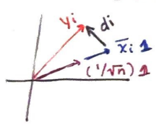
* Definimos el vec $\mathbf{1}_n$ como $\mathbf{1}_n = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \} n$. → long $= \sqrt{n}$

↳ forma un ángulo igual con todos los ejes

↳ el vec $(1/\sqrt{n})\mathbf{1}$ es el vec unitario en dir de $\mathbf{1}_n$.

Si $y_i = \begin{bmatrix} x_{1i} \\ x_{ni} \end{bmatrix}$ → La proyecº ortog. de y_i sobre $\frac{1}{\sqrt{n}}\mathbf{1}$

$[y_i' (\frac{1}{\sqrt{n}}) \mathbf{1}] (\frac{1}{\sqrt{n}} \mathbf{1}) = \left(\frac{x_{1i} + \dots + x_{ni}}{n}\right) \mathbf{1} = \bar{x}_i \mathbf{1}$



$d_i = y_i - \bar{x}_i \mathbf{1}$ es perpendicular a $\bar{x} = \mathbf{1}$

$d_i = y_i - \bar{x}_i \mathbf{1} = \begin{bmatrix} x_{1i} - \bar{x}_i \\ \vdots \\ x_{ni} - \bar{x}_i \end{bmatrix}$

Los elems. de d_i son las desviaciones de la media

$L d_i = \sqrt{d_i' d_i}$

$L^2 d_i = d_i' d_i = \sum^n (x_{ji} - \bar{x}_i)^2 \to \boxed{\frac{L^2 d_i}{n} = s_{ii}}$

⇒ Vecs largos d_i representan más variabilidad en la var i.

$d_i' d_k = \sum (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$

$d_i' d_k = L d_i \, L d_k \cos(\theta_{ik})$

$\sum_j (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$

$\cos(\theta_{ik}) = \frac{s_{ik}}{\sqrt{s_{ii}} \sqrt{s_{kk}}}$ cos del ángulo entre d_i y d_k es la correlaº muestral r_{ik}

Muestras aleatorias y valor esperado de la media muestral y de la matriz de cov. muestral

Supongamos q' NO hemos recogido los datos

$X = \begin{bmatrix} x_{11} & x_{1p} \\ x_{n1} & x_{np} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_n' \end{bmatrix} \}$ vec. aleat.

Los vecs. aleatorios $x_1, \dots, x_n$ son aleatorios indep e idénticamente distrib., si la densidad conjunta es

$f_{x_1}(x_1) \cdot \ldots \cdot f_{x_n}(x_n)$ donde $\underline{f(x_j)} = \underline{f(x_{j1} \dots x_{jp})}$ ⇒ las obs. son indep, pero las variables no necesariamente.

Obs: Las medidas de las p vars. en una misma obs. no necesariam. indep.

Las medidas asociadas a 2 diferentes vars, deben ser indeps.

Supongamos, $X = \begin{bmatrix} x_{11} & x_{1p} \\ x_{n1} & x_{np} \end{bmatrix}$ $\quad y_k = \begin{bmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{bmatrix}$

La ~ de y_k depende de

$f(y_k) = f(x_{1k} \dots x_{nk}) = f_k(x_1) \cdot \ldots \cdot f_k(x_{nk})$

Distribuciones muestrales de $\bar{x}$ y S_n.

Sean $X_1 \dots X_n$ vecs aleat. (muestra multivariada aleat.) tomados de una polaº con media poblacional μ, $[E[X_i] = \mu)$ y cov. poblacional Σ $(var(X_i) = \Sigma)$ $i = 1, \dots, n$

1. $E[\bar{X}] = \mu$

   $cov(\bar{X}) = \frac{1}{n}\Sigma$

2. $E(S_n) = \frac{n-1}{n}\Sigma$

   $E\Big[\underbrace{\frac{n-1}{n} S_n}_{\text{estimador insesgado de } \Sigma}\Big] = \Sigma$

Sesgo de S_n : $B(S_n)$

$B(S_n) = E[S_n] - \Sigma$

$= -\frac{1}{n}\Sigma$

Dem: $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$

$E[\bar{X}] = \frac{1}{n}E[X_1] + \dots + \frac{1}{n}E[X_n]$

$= \underbrace{\frac{1}{n}\mu + \dots + \frac{1}{n}\mu}_{n} = \mu$