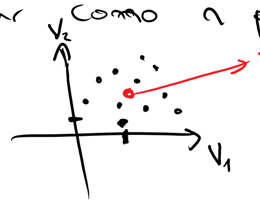
Asumiremos muestras aleatorias.

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix}$$

Primera obs

última obs.

- Cada fila es una observación
- n vectores en $\mathbb{R}^p$.
- Se pueden vizualizar como n puntos en $\mathbb{R}^p$ $(p \leq 3)$



v_1	v_2
3	
2	
1	

$\bar{X} = (\bullet$

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

Una representación geométrica alterna de X es:

* p- vectores en el espacio n-dimensional

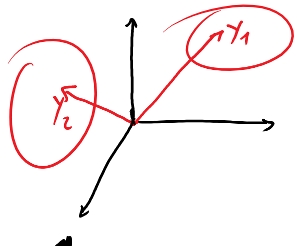
$$\underset{(n \times p)}{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = (Y_1 | \cdots | Y_p)$$

Y_i es un vector de dimensión n $\quad i = 1 \cdots p$

$$Y_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}$$

Son las n mediciones de la variable i $\quad i = 1 \cdots p$.

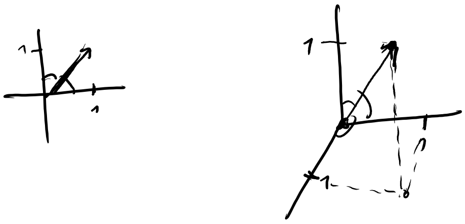
Obs si $n > 3$ no lo podemos visualizar.



Definimos el vector $\mathbf{1}_n$ como:

$$\mathbf{1}_n = \left.\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right\} n$$

Obs: El vector $\mathbf{1}$ forma un ángulo igual con todos los ejes



$$\sqrt{x_1^2 + x_2^2 + \dots x_n^2}$$

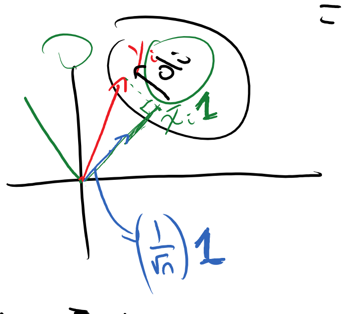
$$\sqrt{n}$$

El vector $\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}$ es el vector unitario en dirección de $\mathbf{1}_n$

Si $y_i = \begin{pmatrix}x_{1i}\\ \vdots\\ x_{ni}\end{pmatrix}$. La proyección ortogonal de y_i sobre $\frac{1}{\sqrt{n}}\mathbf{1}$

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

$$= \bar{x}_i\mathbf{1}$$



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

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

Los elementos de di son las desviaciones de la media

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

$$L^2_{d_i} = d_i' d_i = \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2$$

$$\frac{L^2_{d_i}}{n} = S_{ii} \qquad L^2_{d_i} = n S_{ii}$$

Vectores largos d_i representan más variabilidad en la i-ésima variable.

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

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

$$\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k) = \text{TAREA.}$$

$$\cos(\theta_{ik}) = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = r_{ik}$$

cos del ángulo entre d_i y d_k es el coef. de correlación muestral r_{ik}.

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

Supongamos que NO hemos recogido los datos

$$X = \begin{pmatrix} \overset{\nearrow \text{V.A.}}{X_{11}} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{pmatrix} = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix} \leftarrow \text{vector aleatorio}$$

Los vectores aleatorios $X_1 \dots X_n$ son aleatorios indep e identicamente distribuidos si la densidad conjunta es $\underline{f_{X_1}}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$ donde $= f_X(x)$

$$\underline{f(X_j)} = \underline{f(x_{j1} \dots x_{jp})}$$

Obs: Las medidas de las p variables en una misma obs, no necesariamente indep.

Las medidas asociadas a obs. diferentes deben ser independientes.

Supongamos $X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & & x_{np} \end{pmatrix}$

$$Y_K = \begin{pmatrix} x_{1K} \\ \vdots \\ x_{nK} \end{pmatrix}$$

La distribución de Y_K depende de:

$$f(Y|K) = f(x_{1K} \dots x_{nK}) = f_K(x_{1K}) \cdot f_K(x_{2K}) \cdot \dots \cdot f_K(x_{nK})$$

Distribuciones muestrales de $\bar{X}$ y S_n

Sean $X_1 \dots X_n$ vect. aleatorios (muestra multivariada aleatoria) tomados de una población con media poblacional μ, $\left(E(X_i) = \mu\right)$ y cov. poblacional Σ $\left(Var(X_i) = \Sigma\right)$ $i = 1 \dots n$

Entonces

1) $E(\bar{X}) = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$

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

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

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

Sesgo de S_n

$$B(S_n) = E(S_n) - \Sigma = -\left(\frac{1}{n}\right)\Sigma$$

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

$$E(\bar{X}) = \frac{1}{n}\underbrace{E(X_1)}_{\mu} + \frac{1}{n}\underbrace{E(X_n)}_{\mu} + \dots + \frac{1}{n}\underbrace{E(X_n)}_{\mu}$$

$= \mu$

$E\left((\bar{X}-\mu)(\bar{X}-\mu)'\right)$