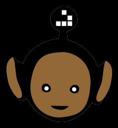
Notas tema introduccion AED

Rodrigo Castillo

5 de septiembre de 2020



1. introduccion

- la mayoria de problemas involucran varias medidas de multiples variables
- extenderemos algunos metodos bla bla bla
- muchos problemas de basan den la distribucion normal multivariable
- sirve mucho

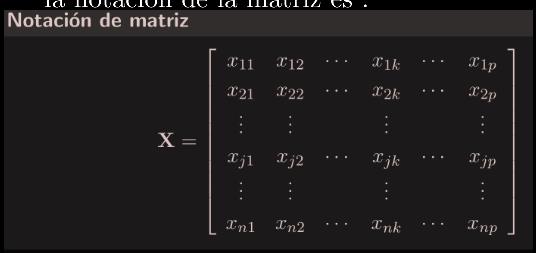
2. Organización y nomenclatura

utilizaremos la notacion x_{ij} como medida de la $k-esima$ variable del $j-esimo$ dato u observacion.

Notación de conjunto de datos

	Variable 1	Variable 2	$\cdots$	Variable k	$\cdots$	Variable p
Item 1:	x_{11}	x_{12}	$\cdots$	x_{1k}	$\cdots$	x_{1p}
Item 2:	x_{21}	x_{22}	$\cdots$	x_{2k}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item j :	x_{j1}	x_{j2}	$\cdots$	x_{jk}	$\cdots$	x_{jp}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item n :	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$\cdots$	x_{np}

la notacion de la matriz es :



Notación de matriz

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}$$

3. Estadistica descriptiva

- generalmente los conjuntos de datos son grandes
- toca crear medidas que resuman los datos

- usualmente se usan medidas que miden la variacion , la ubicacion y la asociacion lineal

Media Muestral:

$$\hat{x_k} = \frac{1}{n}\sum_{j=1}^{n} x_{jk}, k = 1, 2, 3, 4, ..., p \tag{1}$$

Varianza Muestal:

$$s_k^2 = \frac{1}{n}\sum_{j=1}^{n}(x_{jk} - \hat{x_k})^2, k = 1, 2, 3, 4...., p \tag{2}$$

otros datos...

- calcular resumen de los datos
- promedio de los cuadrados de las distancias de todos los numeros con respecto a la media

La varianza se puede considerar como la diagonal de una matriz...

$$s_k^2 = s_{kk} = \frac{1}{n}\sum_{j=1}^{n}(x_{jk} - \hat{x_k})^2, k = 1, 2, 3, 4, 5, ..., p \tag{3}$$

la raiz cuadrada de la varianza muestral $\sqrt{s_{kk}}$ es la desviacion estándar

- la covarianza muestral

$$s_{12} = \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \hat{x_1})(x_{j2} - \hat{x_2}) \tag{4}$$

y de esto se tiene que

- es simetrica si $i = k$

coeficiente de correlacion muestral :

$$r_{ik} = \frac{s_{ik}}{\sqrt{S_{ii}} \times \sqrt{S_{kk}}} \tag{5}$$

media muestral :

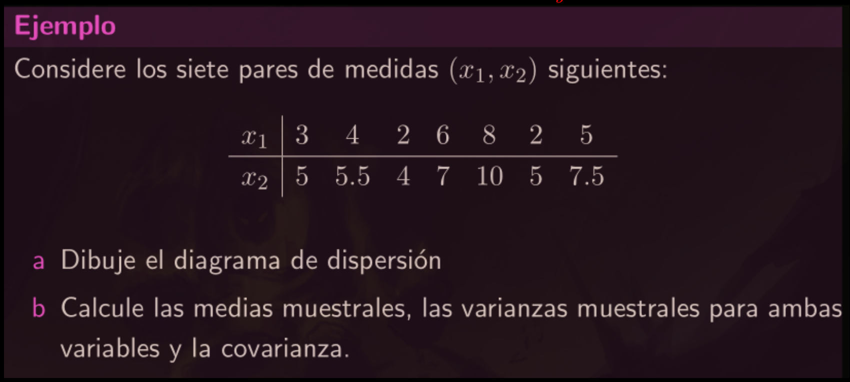
$$\hat{X} = \begin{pmatrix} \hat{x_1} \\ \hat{x_2} \\ . \\ . \\ . \\ \hat{x_p} \end{pmatrix} \tag{6}$$

covarianzas muestrales :

$$S_n = \begin{pmatrix} s_{11} & s_{12} & ... & s_{1p} \\ s_{21} & s_{22} & ... & s_{2p} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ s_{p1} ... & ... & ... & s_{pp} \end{pmatrix} \tag{7}$$

para las Correlaciones muestrales es lo mismo pero con r_{12}, r_{xy}

ejemplo ...



Ejemplo

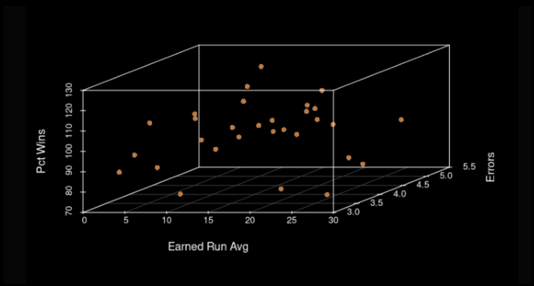
Considere los siete pares de medidas (x_1, x_2) siguientes:

x_1	3	4	2	6	8	2	5
x_2	5	5.5	4	7	10	5	7.5

a Dibuje el diagrama de dispersión

b Calcule las medias muestrales, las varianzas muestrales para ambas variables y la covarianza.

4. Visualizaciones



5. Ejercicios