ESTADÍSTICA

Tema 4: Esperanza y Momentos.

4.4. Covarianza y Correlación.
4.5. Esperanza Condicional.
4.6. Desigualdad de Chebychev.
4.7. Media muestral.

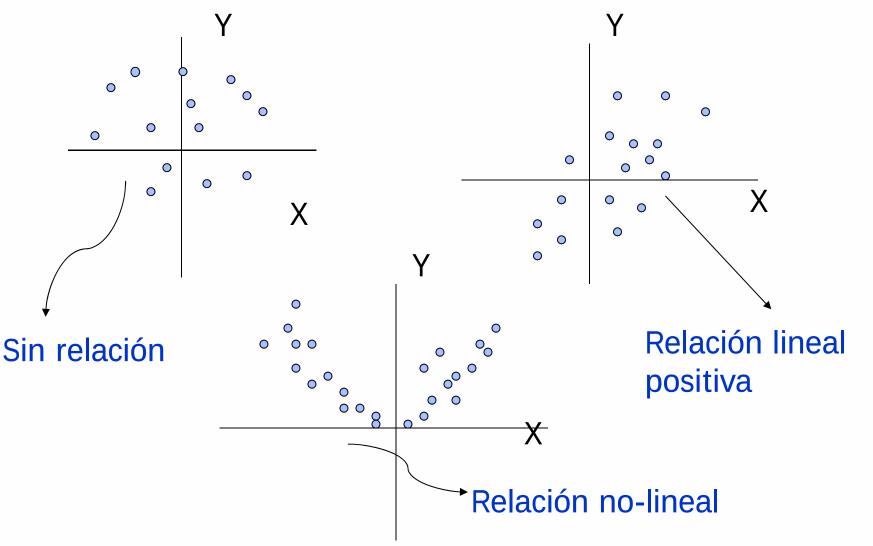
ESTADÍSTICA

Relaciones entre variables (I).

•*Vamos a estudiar a continuación el grado de relación que existe entre dos variables aleatorias.*

•*Para ello, vamos a introducir dos expresiones que miden la relación que existe entre dos variables:* ***la covarianza y la correlación***.

Relaciones entre variables (II).



ESTADÍSTICA

Relaciones entre variables (III).

•*Las relaciones entre v.a. pueden ser de muy distinto tipo: positivas o negativas (si cuando crece la una la otra también lo hace y viceversa), lineales o no lineales, etc.*

•*También puede ocurrir que no exista ninguna relación entre dos v.a.: cuando esto ocurre diremos que* ***dos v.a. son independientes****.*

•*Vamos a estudiar ahora cómo de 'lineal' es la relación que existe entre dos variables: para ello definimos la* ***covarianza*** *y la* ***correlación.***

ESTADÍSTICA

Covarianza y Correlación (I).

***Definición**: Dadas dos variables X e Y, llamaremos covarianza a:*

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$$

- $\mathbf{Cov(X,Y)} = E[(X-E(X))(Y-E(Y))] = E[XY-XE(Y)-YE(X)+E(X)E(Y)] = E(XY)-E(XE(Y))-E(YE(X))+E(E(X)E(Y)) = E(XY)-E(Y)E(X)-E(X)E(Y)+E(X)E(Y) = \mathbf{E(XY)-E(X)E(Y)}$
- *La covarianza puede ser positiva, negativa o nula.*
- *Si las variables X e Y son independientes:* $Cov(X,Y) = E(XY)-E(X)E(Y) = E(X)E(Y)-E(X)E(Y) = 0$
- *Si* $Cov(X,Y)=0$ **NO significa** *que las variables sean independientes.*
- $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

ESTADÍSTICA

Covarianza y Correlación (II).

***Definición**: Dadas dos variables X e Y, llamaremos Coeficiente de Correlación a:*

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- *Su valor está entre* $-1 \leq \rho \leq 1$
- *Si* $\rho = 1$ *ó* $\rho = -1$*, la relación lineal es perfecta* $Y=aX+b$

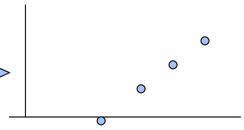
ESTADÍSTICA

ESTADÍSTICA

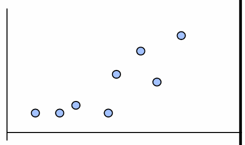
Covarianza y Correlación (III).

Interpretación

$\rho(X,Y)=1$. Existe una relación lineal exacta entre X e Y y la pendiente de la recta es positiva.

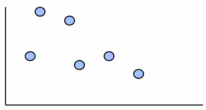


$0<\rho(X,Y)<1$, relación lineal (+) entre X e Y, más intensa cuanto más cercana a 1.

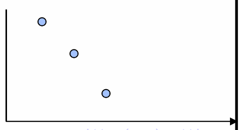


$\rho(X,Y)=0$, ausencia de relación lineal.

$-1<\rho(X,Y)<0$, relación lineal (-) entre X e Y, más intensa cuanto más cercana a -1.



$\rho(X,Y)=-1$. Existe una relación lineal (-) exacta entre X e Y.



ESTADÍSTICA

Esperanza condicional (I).

***Definición**: Dada una v.a. bidimensional (X,Y), se define la esperanza condicional de X (de Y) dado Y=y (dado X=x) como:*

Caso discreto

$$E(X/y)=\sum_x x g_1(x/y)$$

$$E(Y/x)=\sum_y y g_2(y/x)$$

Caso continuo

$$E(X/y)=\int_{-\infty}^{\infty} x\, g_1(x/y)dx$$

$$E(Y/x)=\int_{-\infty}^{\infty} y\, g_2(y/x)dy$$

Esperanza condicional (II).

***Ejemplo**: Dada una v.a. (X,Y), con función de probabilidad:*

Y \ X	2	3	4
2	0	2/4	1/4
1	1/4	0	0

Calcular E(Y/X=2)

$$E(Y/X=2)=\sum_y y g_2(y/x=2)=1g_2(Y=1/X=2)+2g_2(Y=2/X=2)=$$

$$=1\frac{f(2,1)}{f_1(2)}+2\frac{f(2,2)}{f_1(2)}=1\frac{P(X=2,Y=1)}{P(X=2)}+2\frac{P(X=2,Y=2)}{P(X=2)}$$

$$f_1(2)=P(X=2)=\sum_j P(X=2,Y=y_j)=P(X=2,Y=1)+P(X=2,Y=2)=\frac{1}{4}+0=\frac{1}{4}$$

$$E(Y/X=2)=1\frac{P(X=2,Y=1)}{P(X=2)}+2\frac{P(X=2,Y=2)}{P(X=2)}=1\frac{\frac{1}{4}}{\frac{1}{4}}+2\frac{0}{\frac{1}{4}}=1$$

ESTADÍSTICA

Desigualad de Chebychev.

***Desigualdad de Markov.** Sea X una v.a. con esperanza E(X), tal que P(X ≥ 0) = 1. Entonces, dado t > 0*

$$P(X\geq t)\leq\frac{E(X)}{t}$$

***Desigualdad de Chebychev.** Sea X una v.a. con varianza finita y sea t >0. Entonces:*

$$P(|X-E(X)|\geq t)\leq\frac{Var(X)}{t^2}$$

ESTADÍSTICA

Media Muestral.

Definición**: Supongamos que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una variable X con $E(X)=\mu$ y $Var(X)=\sigma^2$. Se define la* ***media muestral *como:*

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Las variables X_i son independientes y todas tienen la misma distribución que X, es decir, $E(X_i)=\mu$ y $Var(X_i)=\sigma^2$

$$E(\overline{X}_n) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{n}(\mu + \ldots + \mu) = \frac{1}{n}n\mu = \mu$$

$$Var(\overline{X}_n) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n^2}Var(X_1 + X_2 + \ldots + X_n) = \frac{1}{n^2}\left[Var(X_1) + \ldots + Var(X_n)\right] = \frac{1}{n^2}(\sigma^2 + \ldots + \sigma^2) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

ESTADÍSTICA

Problema 4.47

Sea X una v.a. que toma valores no negativos y tal que $P(X \geq 6)=1/2$. Probar que $E(X) \geq 3$.
Usar la desigualdad de Markov.

La desiguaaldad de Markov dice:

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Sabemos que $P(X \geq 6)=1/2$, luego $t = 6$

$$P(X \geq 6) \leq \frac{E(X)}{6}$$

$$\frac{1}{2} \leq \frac{E(X)}{6} \quad \Rightarrow \quad \frac{6}{2} \leq E(X) \quad \Rightarrow \quad 3 \leq E(X)$$

ESTADÍSTICA