Educaguía.com

ESTADÍSTICA

DISTRIBUCIONES BIDIMENSIONALES

DISTRIBUCIONES BIDIMENSIONALES

Variables estadísticas bidimensionales

Las variables estadísticas bidimensionales las representaremos por el par(X,Y) donde X es una variable unidimensional que toma los valores $x_1,x_2,x_3,...,x_n$, e Y es otra variable unidimensional que toma los valores $y_1,y_2,y_3,...,y_n$. Por tanto, la variable estadística bidimensional (X,Y) toma estos valores: $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$; o también (x_i,y_i) donde i está comprendido entre los valores 1 y n.

Diagramas de dispersión

Acabamos de ver que los valores de una variable estadística bidimensional son pares de números reales de la forma (xi,yi).
Si representamos estos pares en un sistema de ejes cartesianos se obtiene un conjunto de puntos sobre el plano.
A este conjunto de puntos se le denomina diagrama de dispersión o nube de puntos.

Medias, varianzas y covarianza de una variable estadística bidimensional

Media de la variable X:

$$\bar{x} = \frac{\sum_1^n x_i f_i}{N}$$

Media de la variable Y

$$\bar{y} = \frac{\sum_1^n y_i f_i}{N}$$

Varianza de la variable X

$$s_x^2 = \frac{\sum_1^n x_i^2 f_i}{N} - \bar{x}^2 = \frac{\sum_1^n f_i [x_i - \bar{x}]^2}{N}$$

Varianza de la variable Y

$$s_y^2 = \frac{\sum_1^n y_i^2 f_i}{N} - \bar{x}^2 = \frac{\sum_1^n f_i [y_i - \bar{y}]^2}{N}$$

Covarianza: Se llama covarianza de una variable bidimensional(X,Y) a la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas. La covarianza se representa por s_{xy}:

$$s_{xy} = \frac{\sum_1^n f_i x_i y_i}{N} - \overline{x}\overline{y} = \frac{\sum_1^n f_i (x_i - \overline{x})(y_i - \overline{y})}{N}$$

A la covarianza se le llama tambien varianza conjunta de las variables X e Y.

Concepto general de correlación

Llamamos correlación a la relación o dependencia que existe entre las dos variables que intervienen en la distribución bidimensional.
Hay que tener en cuenta:

- La correlación es lineal o curvilínea según que el diagrama de puntos se condense en torno a una línea recta o a una curva.
- La correlación en positiva o directa cuando a medida que crece una variable la otra también crece.
- La correlación es negativa o inversa cuando a medida que crece una variable la otra decrece.
- La correlación es nula cuando no existe ninguna relación entre ambas variables. En este caso los puntos del diagrama están esparcidos al azar, sin formar ninguna línea, y se dice que las variables están incorreladas.
- La correlación es de tipo funcional si existe una función tal que todos los valores de la distribución la satisfacen. En caso contrario, será tanto más fuerte o más débil dependiendo de la mayor o menor tendencia de los valores de la distribución a satisfacer una determinada función.

Coeficiente de correlación lineal

Una vez que se comprobó con el diagrama de dispersión, la existencia de correlación lineal entre las variables, hay que cuantificar de la forma más precisa posible esta correlación.
La forma mas común de calcular las correlaciones entre variables es el coeficiente de correlación de Pearson.
El coeficiente de correlación lineal de Pearson se define mediante la siguiente expresión:

$$r = \frac{s_{xy}}{s_x s_y}$$

En el que s_{xy} es la covarianza, s_x la desviación típica de la variable X, y s_y la desviación típica de la variable Y.

El signo del coeficiente r viene dado por el signo de la covarianza, ya que las desviaciones típicas son siempre positivas. Así pues, el signo de la covarianza decide el comportamiento de la correlación:

- Si la covarianza es positiva la correlación es directa.
- Si la covarianza es negativa la correlación es inversa.
- Si la covarianza es nula no existe correlación.

Se demuestra que el coeficiente de correlación lineal es un número comprendido entre –1 y 1

Estudio de la dependencia a partir del valor del coeficiente de correlación lineal

Analizando el grado de dependencia entre las variables a partir del coeficiente de correlación lineal, se tiene:

- Si $r = -1$ se puede demostrar que todos los valores de la variable bidimensional se encuentran situados sobre una recta; por tanto satisfacen la ecuación de una recta. Entonces se dice que entre las variables X e Y existe una dependencia funcional.
- Si $-1 < r < 0$ la correlación es negativa y será tanto más fuerte a medida que r se aproxima más a –1 y tanto más débil a medida que se aproxima a 0. En este caso se dice que están en dependencia aleatoria.
- Si $r = 0$ entonces no existe ningún tipo de relación entre las dos variables. En este caso se dice que son aleatoriamente independientes.
- Si $0 < r < 1$ la correlación es positiva y será tanto más fuerte a medida que se aproxime a 1 y tanto más débil a medida que se aproxime a 0. En este caso se dice que están en dependencia aleatoria.
- Si $r = 1$ se puede demostrar que todos los valores de la variable bidimensional se encuentran situados sobre una recta; por tanto satisfacen la ecuación de una recta. Entonces se dice que entre las variables X e Y existe una dependencia funcional.