

Correlación

Dr. Edgar Ramírez Galeano

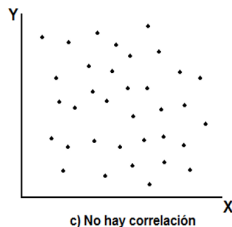
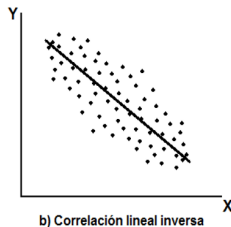
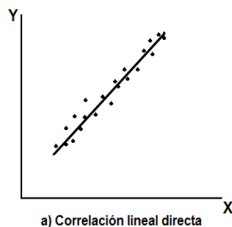
Correlación

Cuando se estudian dos características, una pregunta que surge con frecuencia es si existe alguna relación entre ellas.

Como ejemplo podemos mencionar el peso y la estatura de un grupo de individuos, la temperatura y la tasa de reproducción de una bacteria, el ingreso y el consumo por familias.

CORRELACIÓN LINEAL

Si X e Y son las dos variables a considerar, un diagrama de dispersión muestra la ubicación de los puntos (X, Y) en un sistema de coordenadas rectangular. Si todos los puntos en este diagrama de dispersión se agrupan junto a una recta, como se observa en la Figura 1(a y b), la correlación se llama lineal.



Correlación

La correlación es una medida estadística que indica el grado en que dos variables se relacionan entre sí.

- *Correlación positiva*: las variables aumentan o disminuyen conjuntamente.
- *Correlación negativa*: una variable aumenta mientras la otra disminuye.
- *Inexistente*, cuando no hay relación entre variables.
- Si todos los puntos parecen caer cerca de alguna curva *la correlación se llama no lineal* y la ecuación apropiada no puede ser lineal. Esta correlación también puede ser positiva o negativa.

Propiedades de la Covarianza

La covarianza es una medida de asociación entre dos características que llamaremos X y Y.

$$S_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

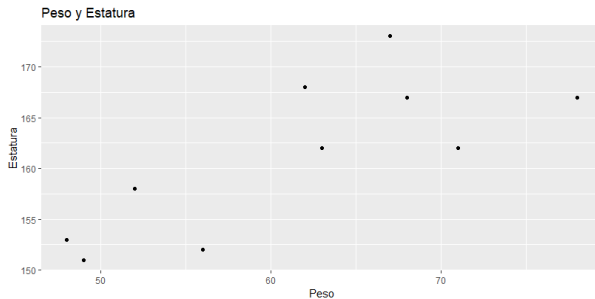
- Si S_{XY} es positivo, relación directa.
- Si S_{XY} es negativo, relación es indirecta.
- Si S_{XY} es cero, no existe relación entre las variables

Ejemplo de Covarianza

Peso	Estatura	Peso	Estatura
63	162	62	168
52	158	68	167
78	167	48	153
49	151	56	152
71	162	67	173

Table: Datos de Peso y Estatura

Ejemplo de Covarianza



x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
63	162	1.6	0.7	1.12
52	158	-9.4	-3.3	31.02
78	167	16.6	5.7	94.62
49	151	-12.4	-10.3	127.72
71	162	9.6	0.7	6.72
62	168	0.6	6.7	4.02
68	167	6.6	5.7	37.62
48	153	-13.4	-8.3	111.22
56	152	-5.4	-9.3	50.22
67	173	5.6	11.7	65.52
614	1613			529.8

$$S_{XY} = \frac{1}{9}(529.80) = 58.86$$

Función en R

```
cov(PesoEstatura$Peso, PesoEstatura$Estatura)
```


Ejemplo de Correlación de Pearson

$$S_{XY} = 58.56$$

$$S_X = 9.98$$

$$S_Y = 7.62$$

$$r_{XY} = \frac{58.56}{(9.98)(7.60)} = 0.7760$$

Función en R

```
cor(PesoEstatura$Peso,PesoEstatura$Estatura)
```

Correlación de Pearson

Es el método de correlación más utilizado pero asume que:

- La tendencia debe ser de tipo lineal.
- No existen valores atípicos (outliers).
- Las variables deben ser numéricas. Si las variables son de tipo ordinal no podremos aplicar la correlación de Pearson.
- Tenemos suficientes datos(algunos autores recomiendan tener más de 30 puntos u observaciones).

Los dos primeros supuestos se pueden evaluar simplemente con un diagrama de dispersión, mientras que para los últimos basta con mirar los datos y el diseño que tenemos.

Para los conjuntos de datos que se describen en cada uno de los retos deberá obtener la Covarianza y el Coeficiente de Correlación.