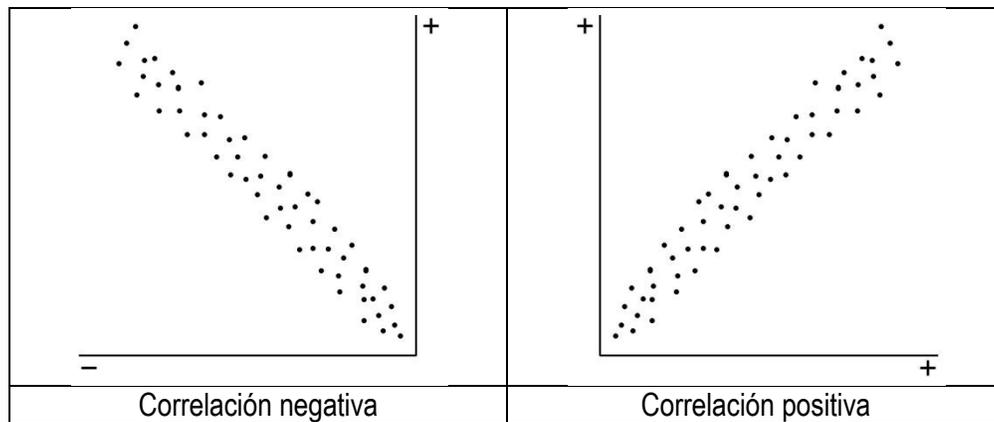


## 5. ANÁLISIS DE CORRELACIONES

Con base en: ROWNTREE, Derek. Introducción a la estadística: un enfoque no matemático. Bogotá: Norma, 1984.

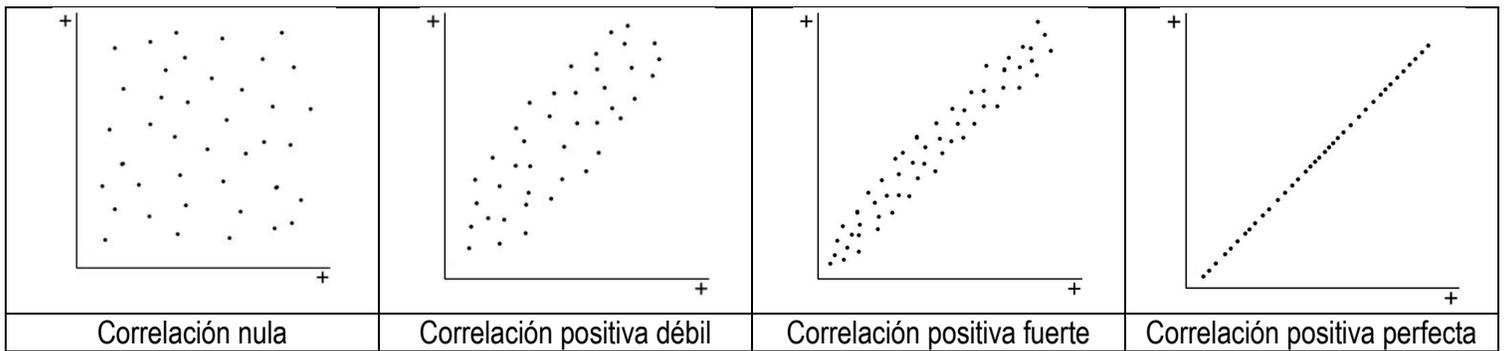
### CORRELACIÓN

Es el estudio de la asociación no causal entre dos (o más) variables, es decir, si los cambios en una variable corresponden a cambios en la otra. Éstos pueden ser en la misma dirección (cuando una aumenta, la otra aumenta: **correlación positiva**) o en dirección contraria (cuando una aumenta, la otra disminuye: **correlación negativa**); o pueden no relacionarse (cuando una aumenta, la otra a veces aumenta, a veces disminuye y a veces sigue igual: **correlación nula**). Se grafica en diagramas de dispersión, como los siguientes.



### Fuerza de la correlación:

Es la cercanía en los cambios de ambas variables. A mayor correlación, más tienden a formar una línea recta en el diagrama de dispersión.



La fuerza de la correlación se mide con:

- Coeficiente de correlación producto-momento (de Pearson): para variables numéricas.
- Coeficiente de correlación de rango (de Spearman): para variable ordinales.

El coeficiente de correlación ( $r$ ) va de **perfecto (-1** ó 1) a nulo (0).

## De la muestra a la población:

¿El  $r$  de la muestra es un estimador confiable del  $r$  de la población? Esto depende de dos factores:

- La magnitud del coeficiente: a más grande, hay menos probabilidad de que se deba a las oscilaciones debidas al azar.
- El tamaño de la muestra: a más grande, más similar será el  $r$  muestral al  $r$  poblacional.

Dado esto, el “Error Estándar del Coeficiente de Correlación” (que es la desviación estándar de una distribución de coeficientes de correlación muestrales) se calcula así:

$$EE_r = \frac{1 - (r)^2}{\sqrt{n}} \quad \text{De forma que el intervalo de confianza al 95\% sería: } r \pm 1.96 EE_r$$

## Significancia del coeficiente:

Si se quiere determinar la seguridad con la cual la relación encontrada es real, y no debida al azar, se aplica una prueba de significancia estadística: prueba  $t$  para coeficientes de correlación. Una forma rápida y **aproximada** es la siguiente:

La hipótesis nula es que  $r = 0$ , entonces: 
$$EE_r = \frac{1 - (0)^2}{\sqrt{n}} = \frac{1}{\sqrt{n}}$$

Así,  $r$  es significativo a un alfa de 0.05, si  $r > 1.96 \frac{1}{\sqrt{n}} = \frac{1.96}{\sqrt{n}}$

## Interpretación del coeficiente:

No hay consenso. Una propuesta común es la siguiente:

Coeficiente	Interpretación
0	Nula
>0.0 – 0.2	Muy baja
>0.2 – 0.4	Baja
>0.4 – 0.6	Moderada
>0.6 – 0.8	Alta
>0.8 – <1.0	Muy alta
1.0	Perfecta

Realmente, si este coeficiente es satisfactorio o no depende del contexto. Ejemplo, los tres casos siguientes:

1. Se encuentra una correlación de +0.65 entre la edad de hombres y mujeres que contraen matrimonio.
2. Se diseña una prueba matemática, pero la critican porque requiere la lectura del planteamiento del problema y esto puede favorecer a las personas con mayor habilidad verbal. Para evaluar esto, se realiza la prueba conjuntamente con otra de habilidad verbal; el coeficiente de correlación entre ambas es cercano a cero.
3. Se propone un nuevo método para medir el porcentaje graso en personas obesas. Se toman dos medidas con el nuevo método y con el antiguo. El coeficiente de correlación entre las dos medidas es +0.92 con el nuevo y +0.98 con el antiguo.

En el primer caso es irrelevante considerar si es satisfactoria o no la correlación obtenida: es simplemente un hecho. En el segundo, la ausencia de correlación es satisfactoria porque indica que la prueba matemática no está sesgada por la habilidad verbal. En el último, el coeficiente de correlación con el nuevo método es insatisfactorio porque el antiguo era más confiable.

Finalmente, se acepta que el cuadrado del coeficiente de correlación indica cuánta variación en una variable puede explicarse por la variación de la otra. Por ejemplo, si la estatura de padres e hijos tiene un  $r$  de 0.8, entonces:  $0.8^2 = 64$ , o sea, el 64% de la variación de la talla de los hijos puede explicarse por la variación de la talla de los padres, y el 36% por otros factores.