

# UNIDAD 3: REGRESIÓN Y CORRELACIÓN

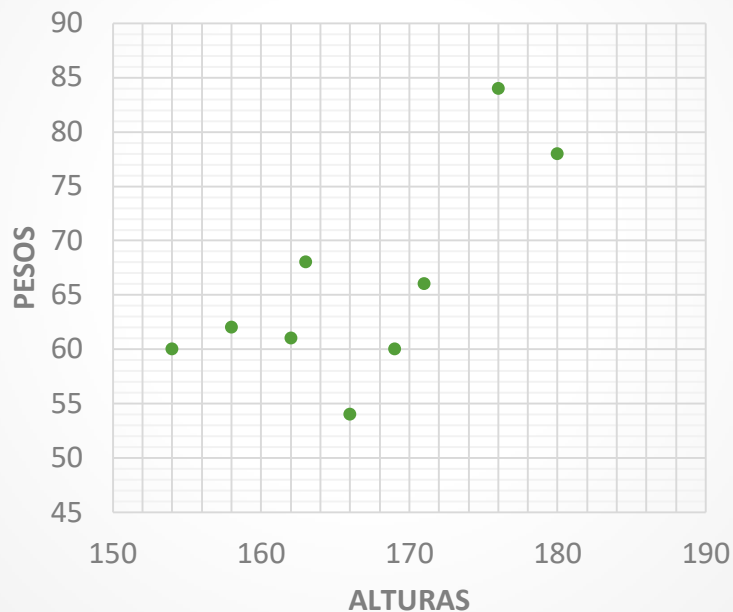
---

ESTADÍSTICA APLICADA

2020

# ANÁLISIS DE REGRESIÓN

ALTURA Y PESO DE INDIVIDUOS



Altura en cm	Peso en Kg
162	61
154	60
180	78
158	62
171	66
169	60
166	54
176	84
163	68

ESTUDIA LA RELACIÓN FUNCIONAL QUE EXISTE ENTRE DOS O MÁS VARIABLES CUANTITATIVAS.

A partir de los siguientes datos obtenidos, se observa dos variables en varios individuos de una muestra.

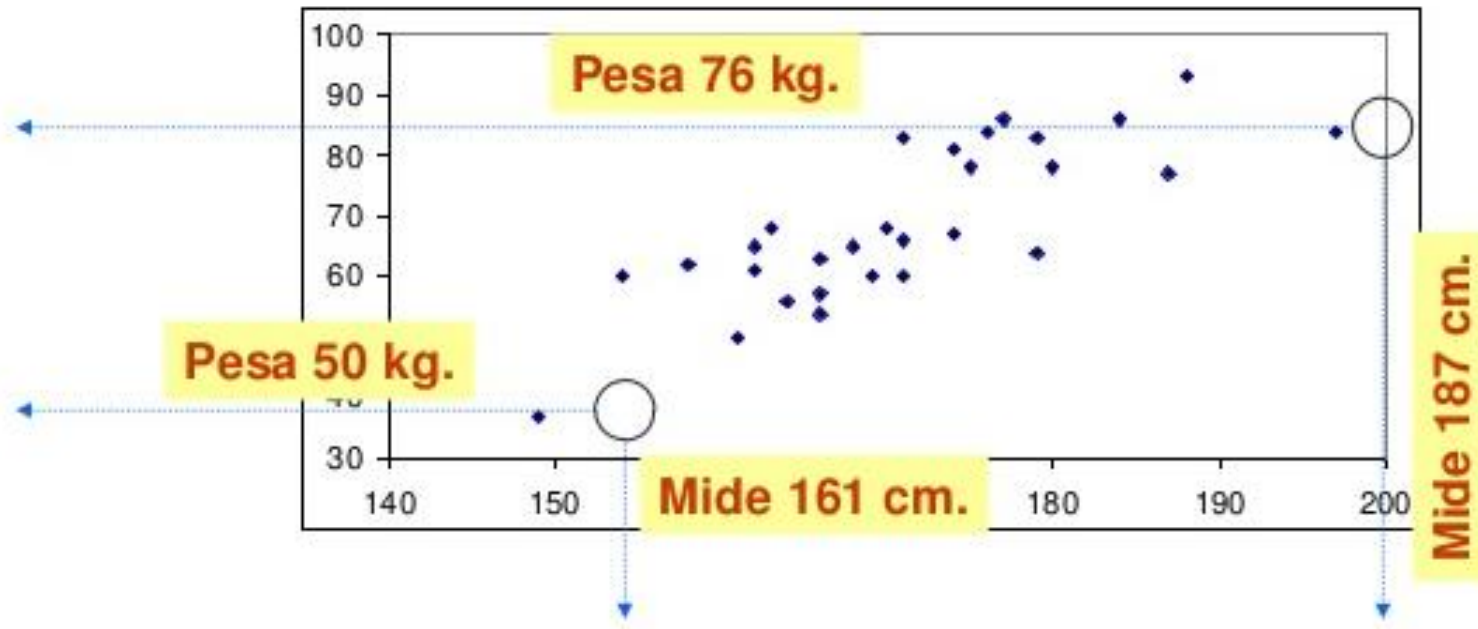
En cada fila tenemos los datos de un individuo y cada columna representa los valores que toma una variable sobre los mismos.

Dichas observaciones pueden ser representadas en un diagrama de dispersión. En ellos cada individuo es un punto, cuyas coordenadas son los valores de la variable.

Nuestro objetivo será **reconocer** a partir de los mismos si hay **relación** entre las variables, de que **tipo** es y si es posible **predecir** el valor de una de ellas en función de la otra.

# DIAGRAMA DE DISPERSIÓN O NUBES DE PUNTOS

---



# ANÁLISIS DE REGRESIÓN

---

POR LO TANTO EN EL ANÁLISIS DE REGRESIÓN NOS INTERESA:

- ✓ Investigar si existe una asociación entre las dos variables testeando la hipótesis de independencia estadística.
- ✓ Estudiar la fuerza de la asociación, a través de una medida de asociación denominada coeficiente de correlación.
- ✓ Estudiar la forma de la relación. Usando los datos propondremos un modelo para la relación y a partir de ella será posible predecir el valor de una variable a partir de la otra.

# ANÁLISIS DE CORRELACIÓN

---

**El Análisis de Correlación lineal estudia el grado y sentido de la asociación lineal que hay entre un conjunto de variables**  $y$ , a diferencia del análisis de regresión, no se identifica ni se estima explícitamente un modelo funcional para las variables, este siempre se supone lineal. El interés principal es medir la asociación entre dos variables aleatorias cualesquiera, sin necesidad de distinguir variables dependientes e independientes.

En el análisis de correlación, ninguna de las variables puede ser fijada por el experimentador, ya que éste podría seleccionar niveles de las variables que no son frecuentes y esto podría conducir a una estimación errada del grado de correlación.

# RECTA DE REGRESIÓN

Generalmente, en los casos de correlación lineal, los puntos que representan los pares ordenados de valores de las variables (x,y), no están situados sobre la línea recta (correlación perfecta), sino que se encuentran entorno a ella.

El problema que se plantea es el de determinar la ecuación de la recta mas representativa del conjunto de puntos, es decir, la recta con respecto a la cual la dispersión de los puntos es mínima.

Esta recta se llama recta de regresión.

La ecuación de la recta de regresión es: 
$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$



# COEFICIENTE DE CORRELACIÓN LINEAL

---

La simple observación de que dos variables parecen estar relacionadas, no revela gran cosa. Dos importantes preguntas se pueden formular al respecto: **¿Qué tan estrechamente relacionadas se encuentran las variables?** o **¿cuál es el grado de asociación que existe entre ambas?**

Para responder a la primer pregunta se necesita una medida del **grado de asociación entre las dos variables**. Esta medida es el **coeficiente de correlación**, que se denota con la letra **r**.

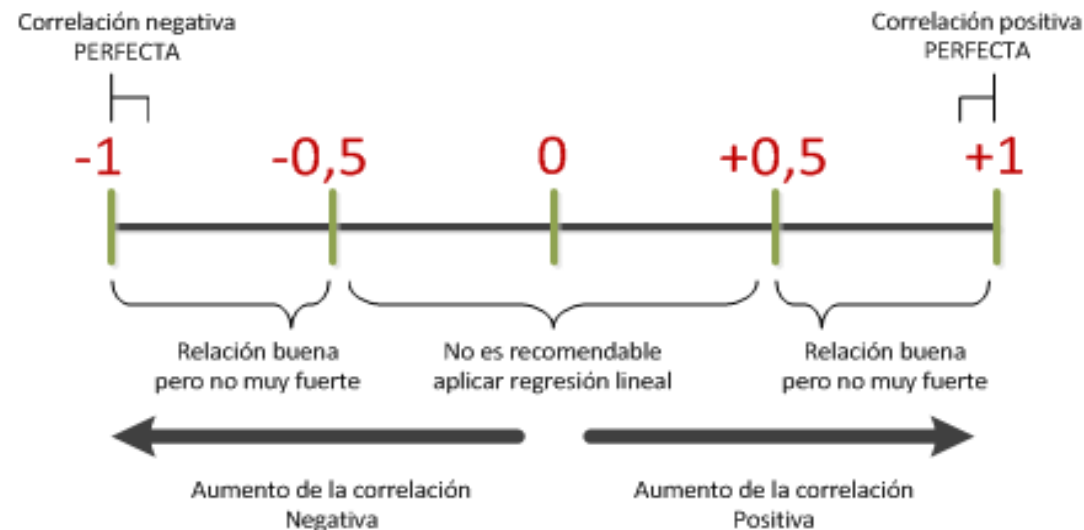
El coeficiente de correlación lineal es el cociente entre la covarianza y el producto de las desviaciones típicas de ambas variables.

# COEFICIENTE DE CORRELACIÓN

Para medir la correlación hay dos valores extremos: 1 y -1. por consiguiente a cada valor constante comprendido en el intervalo [-1,1], llamado coeficiente de correlación lineal.

Para determinar el **coeficiente de correlación** se ha establecido la siguiente expresión:

$$r = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$





# COEFICIENTE DE CORRELACIÓN

---

Hablaremos de correlación lineal fuerte cuando la nube se parezca mucho a una recta y será cada vez más débil (o menos fuerte) cuando la nube vaya desparramándose con respecto a la recta.

En el gráfico observamos que en nuestro ejemplo la correlación es bastante fuerte, ya que la recta que hemos dibujado está próxima a los puntos de la nube.

Cuando la recta es creciente la correlación es positiva o directa: al aumentar una variable, la otra tiene también tendencia a aumentar, como en el ejemplo anterior. Cuando la recta es decreciente la correlación es negativa o inversa: al aumentar una variable, la otra tiene tendencia a disminuir.

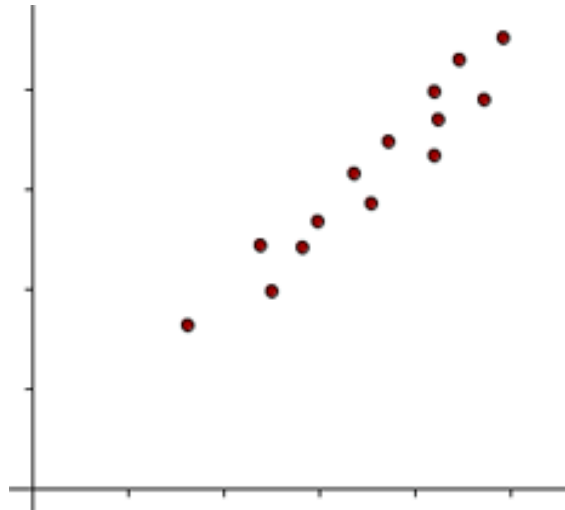
# GRADO DE CORRELACIÓN

---

El grado de correlación indica la proximidad que hay entre los puntos de la nube de puntos. Se pueden dar tres tipos:

## **CORRELACIÓN FUERTE**

La correlación será fuerte cuanto más cerca estén los puntos de la recta.

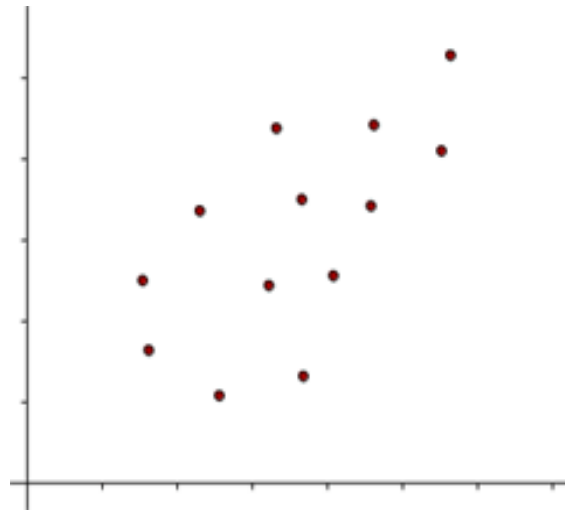


# GRADO DE CORRELACIÓN

---

## CORRELACIÓN DÉBIL

La correlación será débil cuanto más separados estén los puntos de la recta.



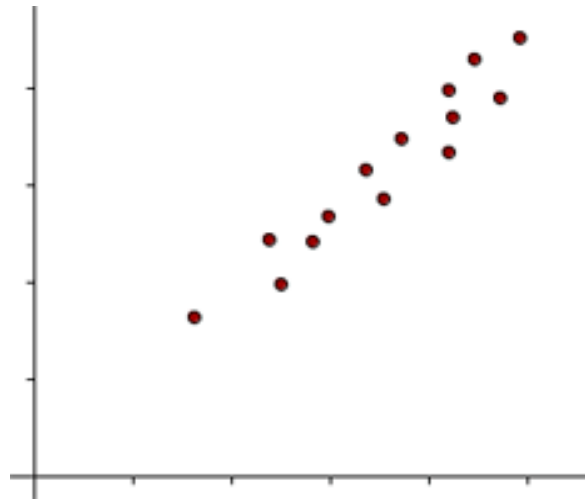
# TIPOS DE CORRELACIÓN

---

## **CORRELACIÓN DIRECTA**

La correlación directa se da cuando al aumentar una de las variables la otra aumenta.

La recta correspondiente a la nube de puntos de la distribución es una recta creciente.



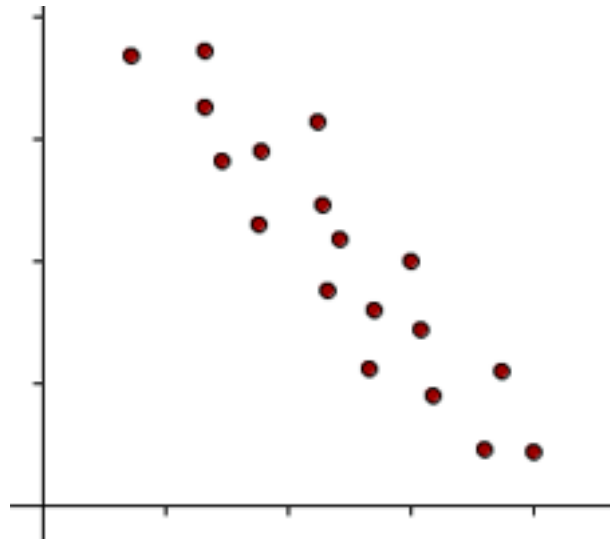
# TIPOS DE CORRELACIÓN

---

## **CORRELACIÓN INVERSA:**

La correlación inversa se da cuando al aumentar una de las variables la otra disminuye.

La recta correspondiente a la nube de puntos de la distribución es una recta decreciente.



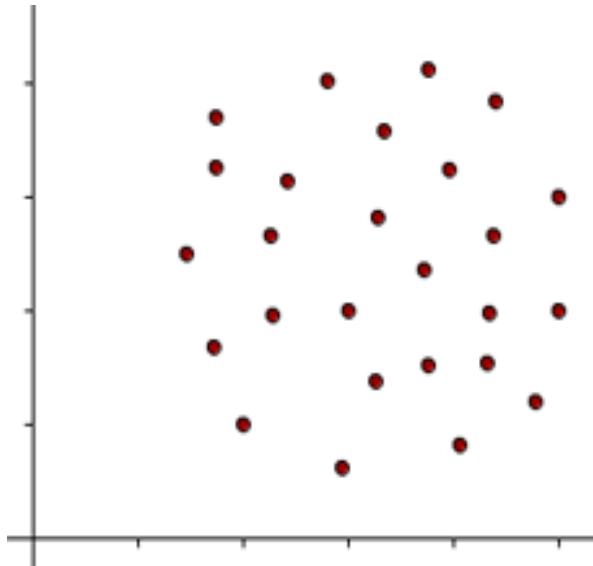
# TIPOS DE CORRELACIÓN

---

## **CORRELACIÓN NULA**

La correlación nula se da cuando no hay dependencia de ningún tipo entre las variables.

En este caso se dice que las variables son incorreladas y la nube de puntos tiene una forma redondeada.



# PROPIEDADES DEL COEFICIENTE DE CORRELACIÓN

---

- ✓ El coeficiente de correlación no varía al hacerlo la escala de medición.

Es decir, si expresamos la altura en metros o en centímetros el coeficiente de correlación no varía.

- ✓ El signo del coeficiente de correlación es el mismo que el de la covarianza.

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.

- ✓ El coeficiente de correlación lineal es un número real comprendido entre  $-1$  y  $1$ .

- ✓ Si el coeficiente de correlación lineal toma valores cercanos a  $-1$  la correlación es fuerte e inversa, y será tanto más fuerte cuanto más se aproxime  $r$  a  $-1$ .

- ✓ Si el coeficiente de correlación lineal toma valores cercanos a  $1$  la correlación es fuerte y directa, y será tanto más fuerte cuanto más se aproxime  $r$  a  $1$ .

- ✓ Si el coeficiente de correlación lineal toma valores cercanos a  $0$ , la correlación es débil.

- ✓ Si  $r = 1$  o  $-1$ , los puntos de la nube están sobre la recta creciente o decreciente. Entre ambas variables hay dependencia funcional.

# RECTA DE REGRESIÓN

---

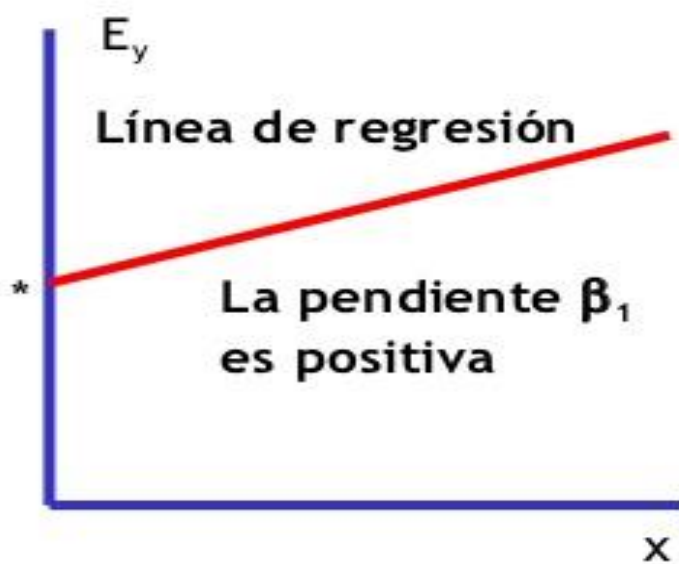
- Es la que mejor se ajusta a la nube de puntos.
- Pasa por el punto llamado centro de gravedad  $(\bar{x}; \bar{y})$ .
- Se utiliza para estimar los valores de la Y a partir de los de la X.
- La pendiente de la recta es el cociente entre la covarianza y la varianza de la variable X.
- Se utiliza para estimar los valores de la X a partir de los de Y.
- La pendiente de la recta es el cociente entre la covarianza y la varianza de la variable Y.



# Líneas posibles de regresión en la regresión lineal simple

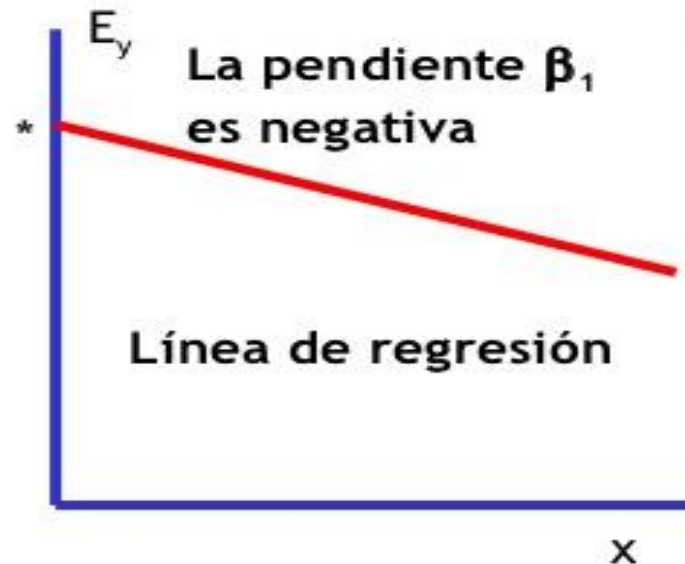
## Sección A

Relación lineal positiva



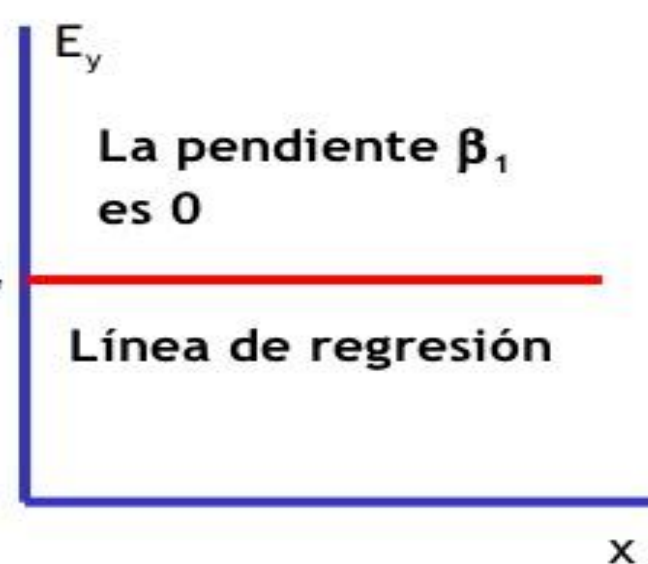
## Sección B

Relación lineal negativa



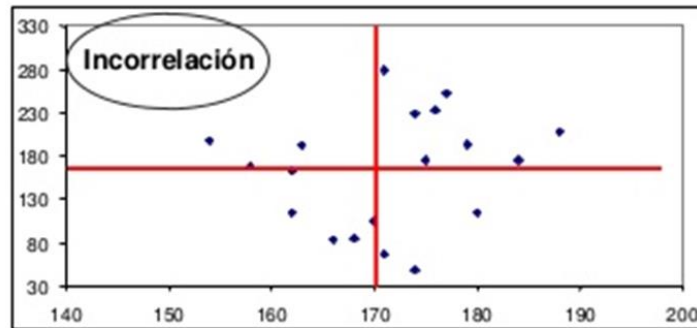
## Sección C

No hay relación



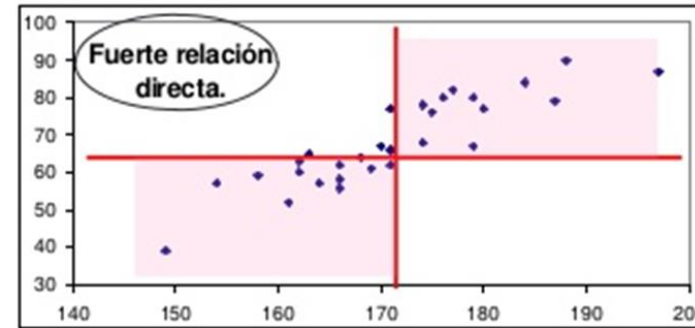
\* Ordenada al origen  $\beta_0$

# CORRELACIÓN DIRECTA E INVERSA



Para valores de X por encima de la media tenemos valores de Y por encima y por debajo en proporciones similares.

**Incorrelación.**

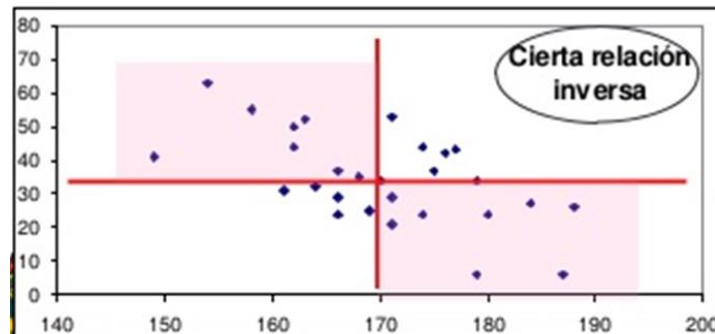


- Para los valores de X mayores que la media le corresponden valores de Y mayores también.

- Para los valores de X menores que la media le corresponden valores de Y menores también.

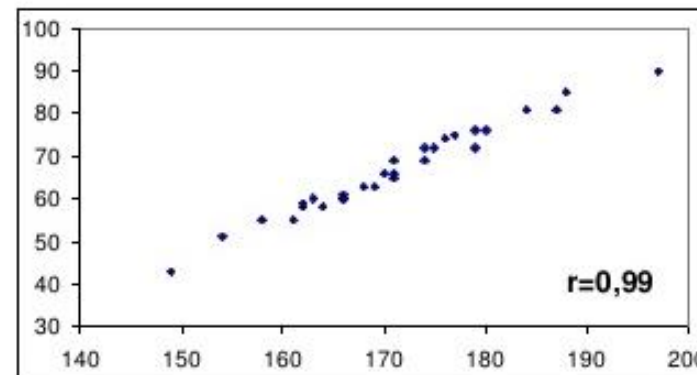
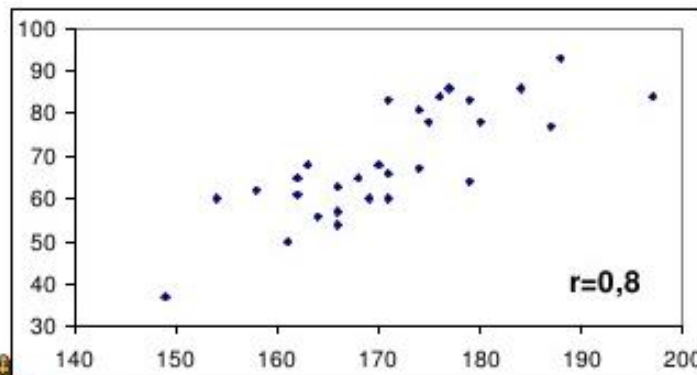
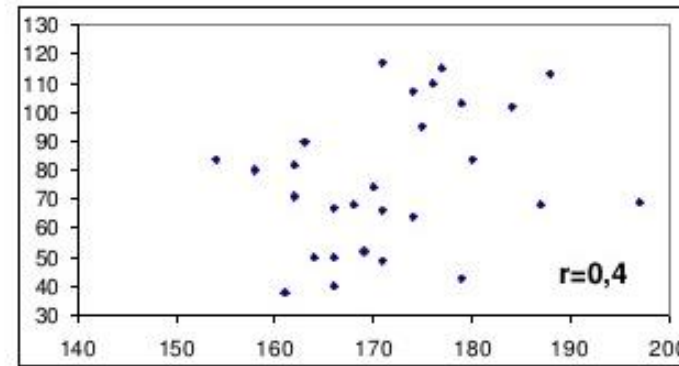
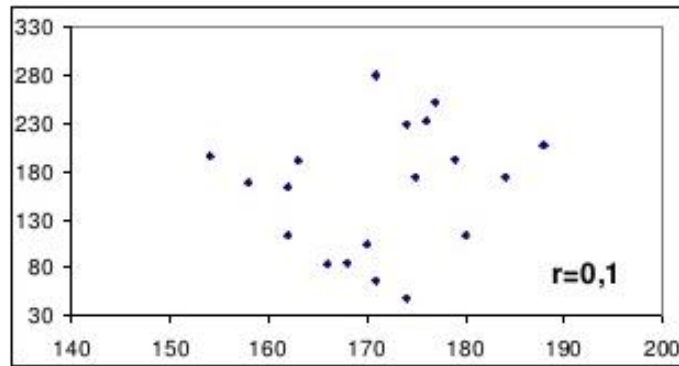
- Esto se llama **relación directa.**

Para los valores de X mayores que la media le corresponden valores de Y menores. Esto es **relación inversa** o decreciente.



# CORRELACIONES POSITIVAS

## Entrenando el ojo: correlaciones positivas



# CORRELACIONES NEGATIVAS

## Entrenando el ojo: correlaciones negativas

