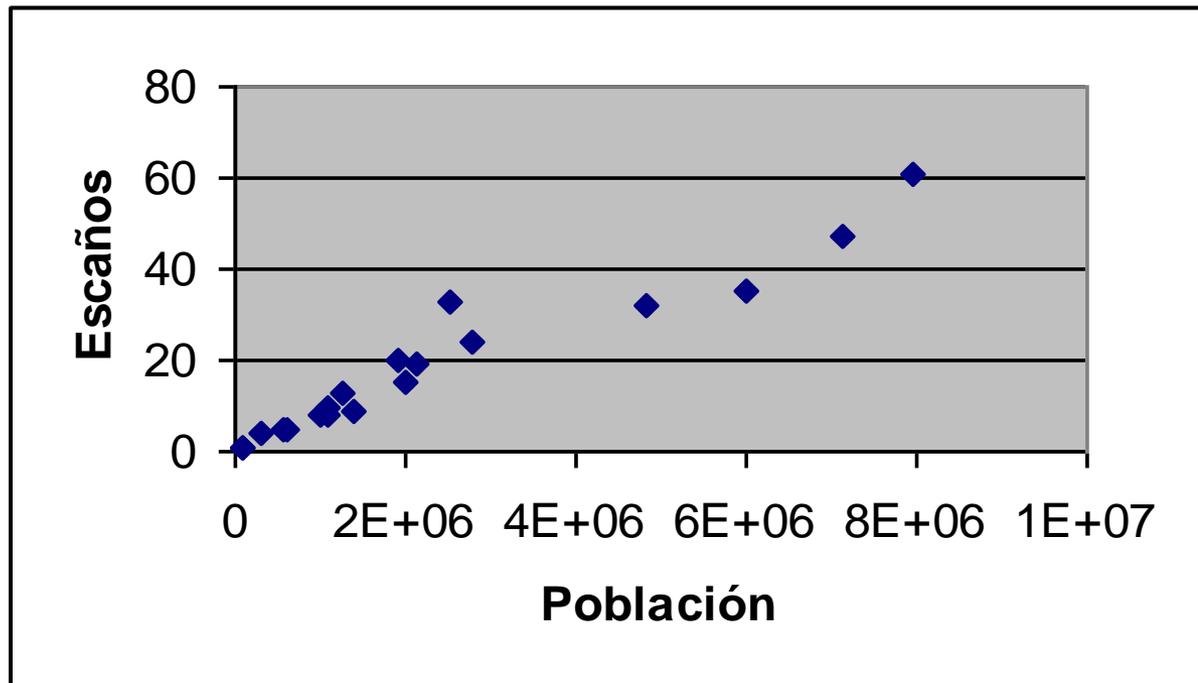




3.2: Medidas numéricas



Hemos visto que había una relación aproximadamente lineal entre población y esaños. Buscamos una medida de la fuerza de la relación



La covarianza

- La **covarianza** mide la fuerza de la relación lineal entre **dos variables**
- La **covarianza muestral** puede calcularse mediante:

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

– Una alta covarianza no implica efecto causal



Una fórmula alternativa para la covarianza

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x} \bar{y}$$

Si tenemos que calcular la covarianza a mano, esta fórmula es más fácil.



Interpretación de la covarianza

La **covarianza** entre dos variables:

$\text{Cov}(x,y) > 0$: X e Y tienden a moverse en la misma dirección

$\text{Cov}(x,y) < 0$: X e Y tienden a moverse en direcciones opuestas.

$\text{Cov}(x,y) = 0$: X e Y no están relacionadas linealmente.



Inconveniente de la covarianza

En nuestro ejemplo, la covarianza es aproximadamente 36043027,5. ¿Indica una relación fuerte o no?

¿Cuáles son las unidades de la covarianza?

¿Cómo podemos corregir el problema?



La correlación

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

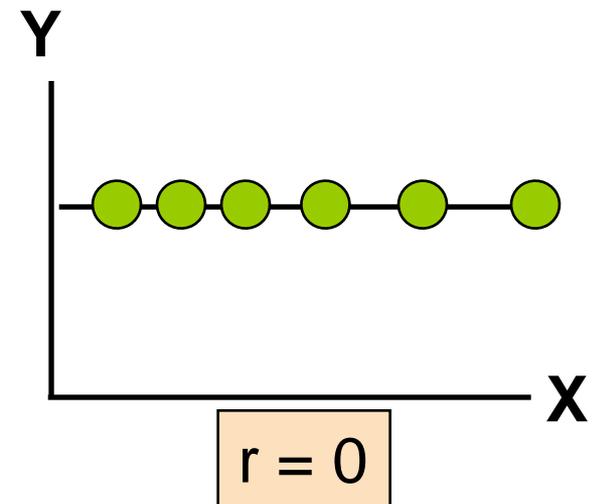
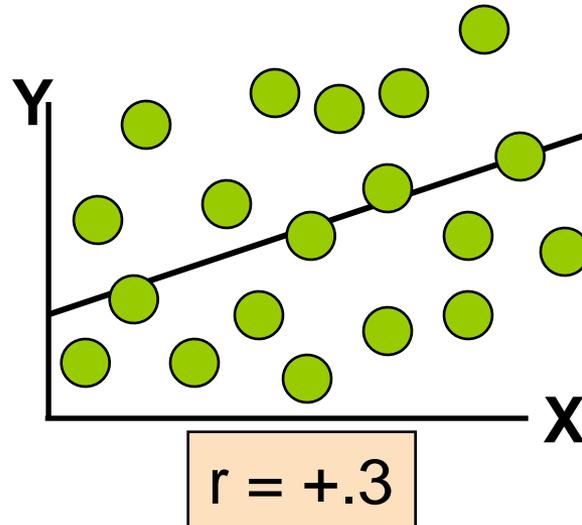
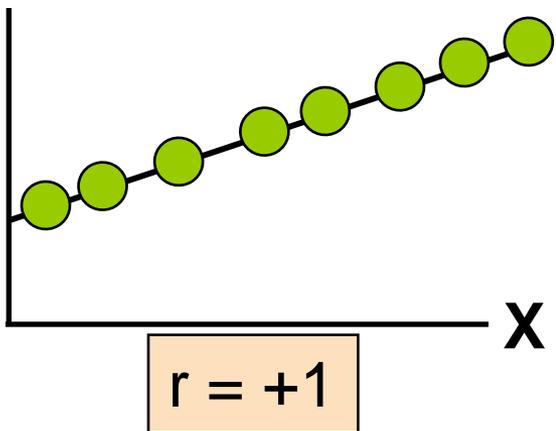
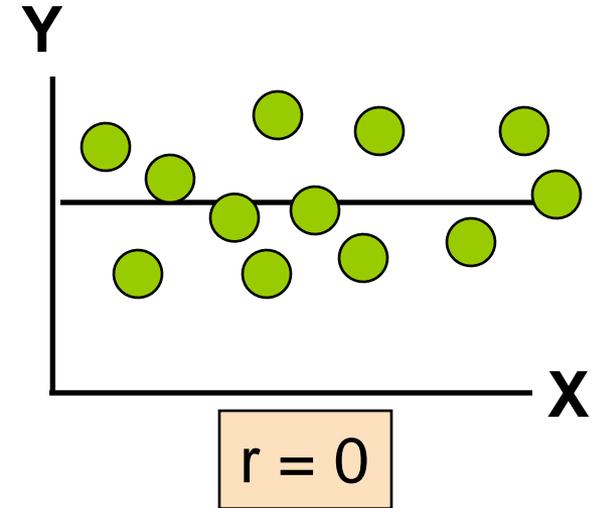
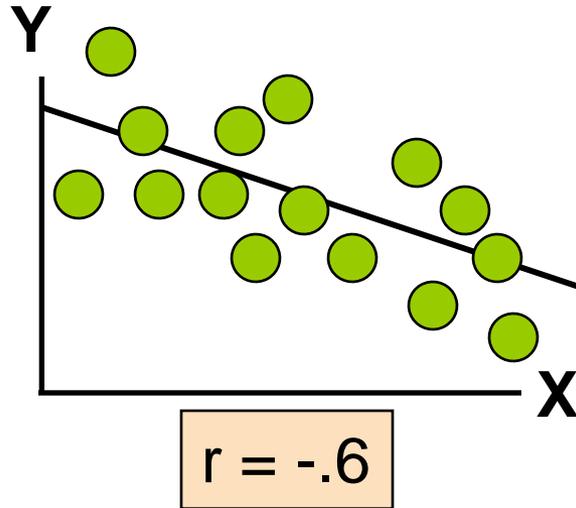
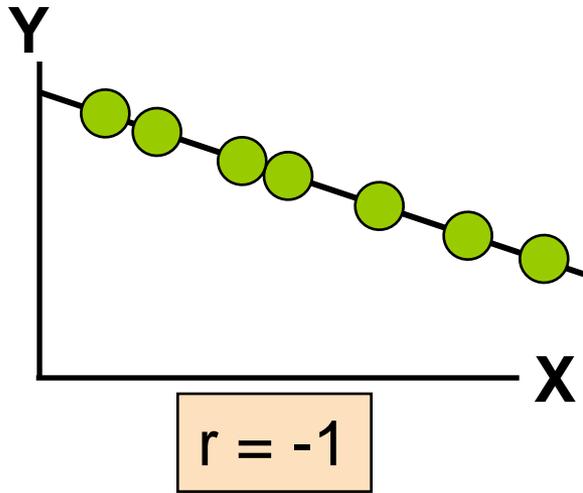
$$-1 \leq r \leq 1$$

$r = 1$: hay una relación positiva perfecta

$r = -1$: hay una relación lineal negativa perfecta

$r = 0$: no existe relación lineal, datos incorreladas

En el ejemplo, $r=0,967$: una relación fuerte y positiva





Cálculo de la covarianza y correlación mediante la tabla de frecuencias conjuntas

Los siguientes datos son resultados de una encuesta de alumnos de políticas sobre la asignatura de estadística.

		Cantidad de trabajo hecho					Total
		1	2	3	4	5	
Satisfacción con el profesor	1	2	2	0	1	2	7
	2	5	3	1	2	3	14
	3	2	2	8	1	3	16
	4	1	2	4	6	2	15
	5	0	1	4	7	8	20
Total		10	10	17	17	18	72



Calculamos la tabla de frecuencias relativas ...

		Cantidad de trabajo hecho					Total
		1	2	3	4	5	
Satisfacción	1	0,028	0,028	0,000	0,014	0,028	0,097
	2	0,069	0,042	0,014	0,028	0,042	0,194
	3	0,028	0,028	0,111	0,014	0,042	0,222
	4	0,014	0,028	0,056	0,083	0,028	0,208
	5	0,000	0,014	0,056	0,097	0,111	0,278
	Total		0,139	0,139	0,236	0,236	0,250

y las medias marginales

	1	2	3	4	5
0,139	0,139	0,139	0,236	0,236	0,250
0,139	0,139	0,278	0,708	0,944	1,250

3,319

1	0,097	0,097
2	0,194	0,389
3	0,222	0,667
4	0,208	0,833
5	0,278	1,389

3,375



Ahora calculamos la suma de frecuencia relativa x satisfacción x trabajo.

		Cantidad de trabajo hecho				
		1	2	3	4	5
Satisfacción	1	0,028	0,056	0,000	0,056	0,139
	2	0,139	0,167	0,083	0,222	0,417
	3	0,083	0,167	1,000	0,167	0,625
	4	0,056	0,222	0,667	1,333	0,556
	5	0,000	0,139	0,833	1,944	2,778
					11,880	

Luego la covarianza es $11,880 - 3,319 \times 3,375 = 0,672$

Hay una relación positiva entre las dos variables.



Para calcular la correlación, primero calculamos las varianzas ...

1	2	3	4	5
0,139	0,139	0,236	0,236	0,250
0,139	0,556	2,125	3,778	6,250

12,85

La varianza es $12,85 - 3,319^2 = 1,829$

1	0,097	0,097
2	0,194	0,778
3	0,222	2,000
4	0,208	3,333
5	0,278	6,944

13,153

La varianza es $13,153 - 3,375^2 = 1,762$



Finalmente dividimos la covarianza por el producto de las raíces de las varianzas para sacar la correlación.

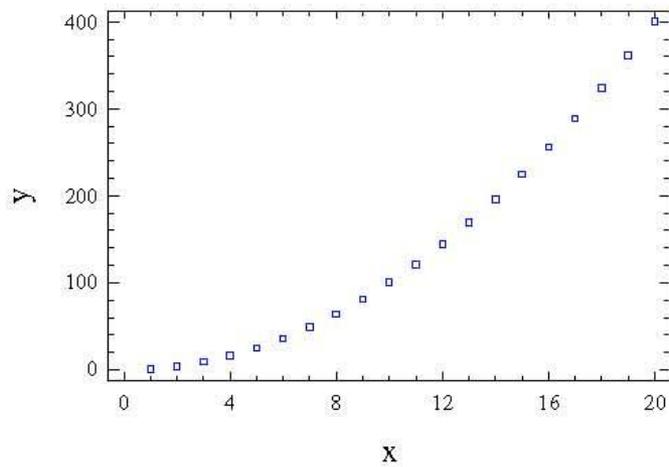
Nos sale el valor de **0,374**.

Una ligera relación positiva entre las dos variables.

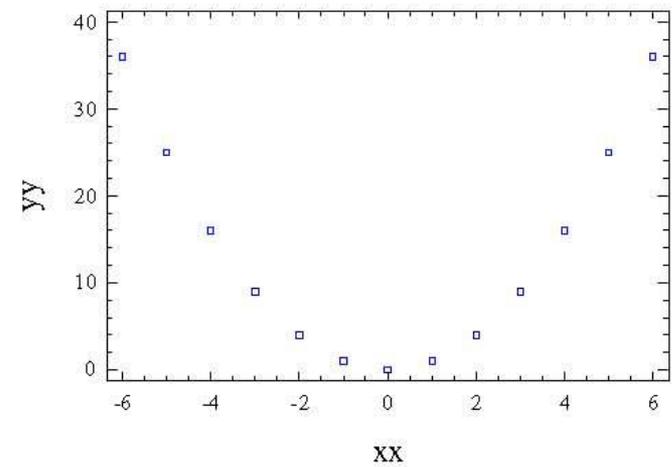


Correlación y relaciones no lineales

Correlación = 0.97



Correlación = 0



En ambos gráficos se ha usado la relación $y=x^2$. ¡Una fuerte relación no lineal!



Correlación y causalidad I





Correlación y causalidad II

Homero: *No hay siquiera un oso a la vista. ¡La "patrulla anti-osos" funciona de maravilla!*

Lisa: *Eso es un razonamiento falaz, Papá.*

Homero [sin comprender]: *Gracias, hija.*

Lisa: *Usando tu lógica, yo puedo afirmar que esta roca aleja a los tigres.*

Homero: *Hmmm, ¿y cómo funciona?*

Lisa: *No funciona. (pausa) ¡Es sólo una roca estúpida!*

Homero: *Ajá.*

Lisa: *Pero no veo ningún tigre alrededor, ¿y tú?*

Homero: *(. . . pausa . . .) Lisa, quiero comprar tu roca.*



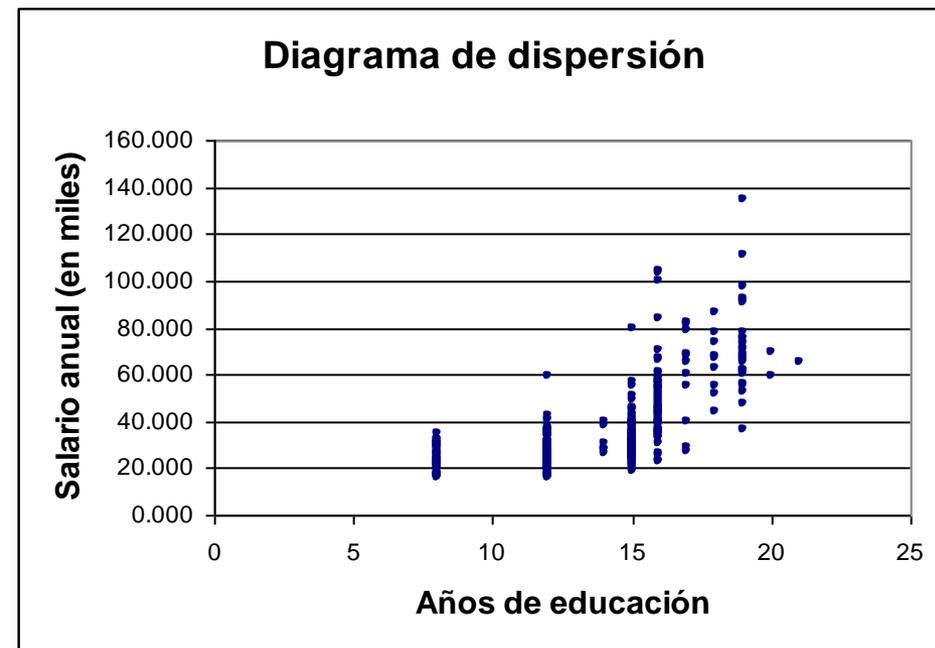


Ejercicio (Pregunta de Test)

Se ha realizado una encuesta a 474 empleados de una compañía multinacional. Entre los datos recogidos consta el salario anual (en miles) y los años de educación. Al realizar el diagrama de dispersión asumiendo que el salario depende de los años de educación se observa la siguiente nube de puntos:

Señala cual de las siguientes opciones es la correcta:

- a) La covarianza debe ser positiva y la correlación negativa.
- b) La covarianza debe ser positiva y la correlación positiva.
- c) La covarianza debe ser negativa y la correlación negativa.
- d) La covarianza debe ser negativa y la correlación positiva.





Ejercicio (Pregunta de Test)

Se ha realizado una encuesta a 474 empleados de una compañía multinacional. Entre los datos recogidos consta el salario anual (en miles) y los años de educación. Suponiendo Y =Salario, X =Años de educación

$$\text{Varianza } X = 8,305 \quad \text{Varianza } Y = 290,963 \quad \text{Covarianza} = 32,471$$

Señala cual es el valor correcto de la correlación:

- a) -0,53
- b) 0,066
- c) -0,662
- d) 0,662