

Estadística aplicada a la comunicación

Tema 5: Análisis de datos cuantitativos I: estadística descriptiva

b. Análisis bivalente

OpenCourseWare UPV/EHU

Unai Martín Roncero

Departamento de Sociología 2

unai.martin@ehu.eus

OCW
OpenCourseWare



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Estadística bivalente

5.2.1 Introducción

5.2.2 Asociación entre variables cuantitativas

5.2.3 Asociación y dependencia entre variables cualitativas

5.2.1 Introduction

Objetivo: determinar si existe relación entre dos variables:

- ¿Influye el sexo en el tipo de coche preferido?
- ¿Influye la renta en el consumo de alcohol?
- ...

Estadística inferencial y teoría de la probabilidad

- ¿Es esa relación extrapolable a la población?

5.2.1 Introduction

El tipo de variable determina el procedimiento a utilizar:

- Dos variables cualitativas: ¿Influye el sexo en la elección de carrera?
 - ✓ Técnicas de asociación: tablas de contingencia, chi-cuadrado
- Dos variables cuantitativas: ¿Hay relación entre la edad y los ingresos?
 - ✓ Coeficientes de correlación: Pearson
- Una variable cualitativa y otra cuantitativa: ¿Vivir en un pueblo influye en la valoración de un partido político?
 - ✓ Análisis de la varianza

5.2.1 Introduction

Dos conceptos importantes:

- Variable dependiente (respuesta) vs. Variable independiente (predictora)
- Variable confusora: es una variable que se relaciona (directa o inversamente) con la variable dependiente e independiente y confunde la relación entre ambas.
 - ✓ ¿Influye el sexo en el tipo de coche preferido?
 - ✓ ¿Influye la renta en el consumo de alcohol?
 - ✓ ¿Influye el consumo de café en el cáncer de pulmón?

5.2.2 Asociación entre variables cuantitativas

Para estudiar la relación entre dos variables cuantitativas contamos con:

- Gráficos de dispersión
- Coeficientes de correlación

5.2.2 Asociación entre variables cuantitativas

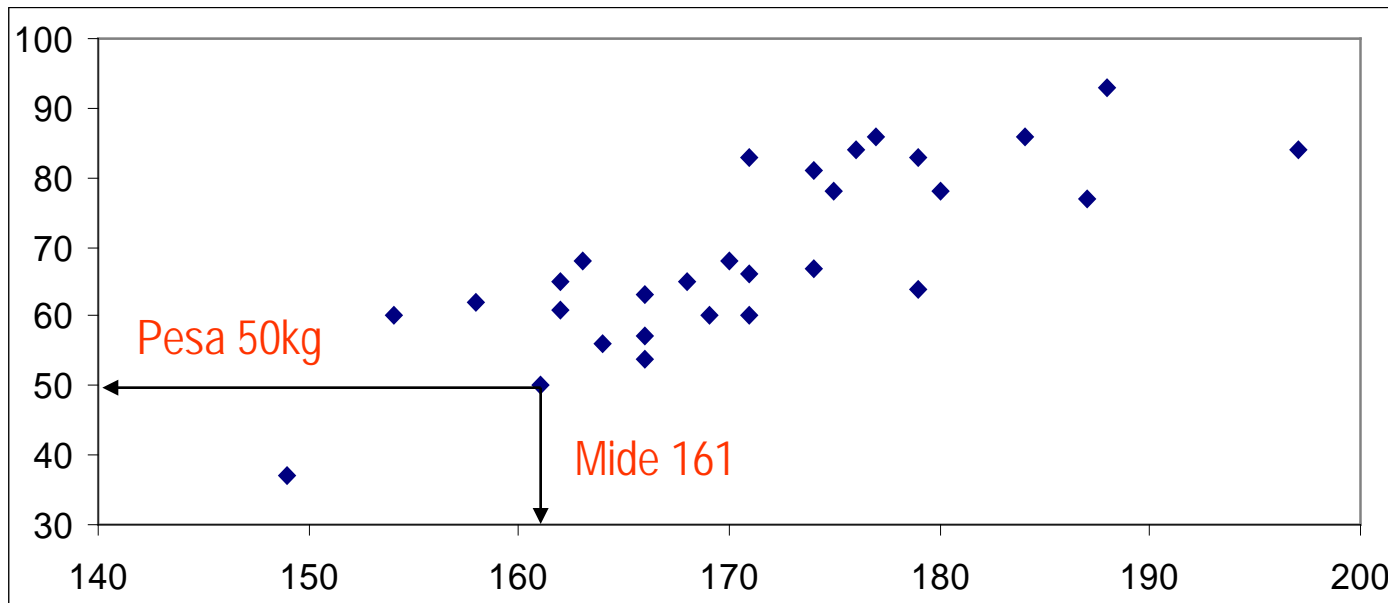
¿Cómo medimos la relación lineal entre dos variables? (Ejemplo tomado de Barón 2004)

Tenemos las alturas y los pesos de 30 individuos y deseamos conocer si existe relación entre la altura y el peso, es decir, saber si la altura de una persona influye en su peso y en el caso de que exista relación lineal de qué tipo.

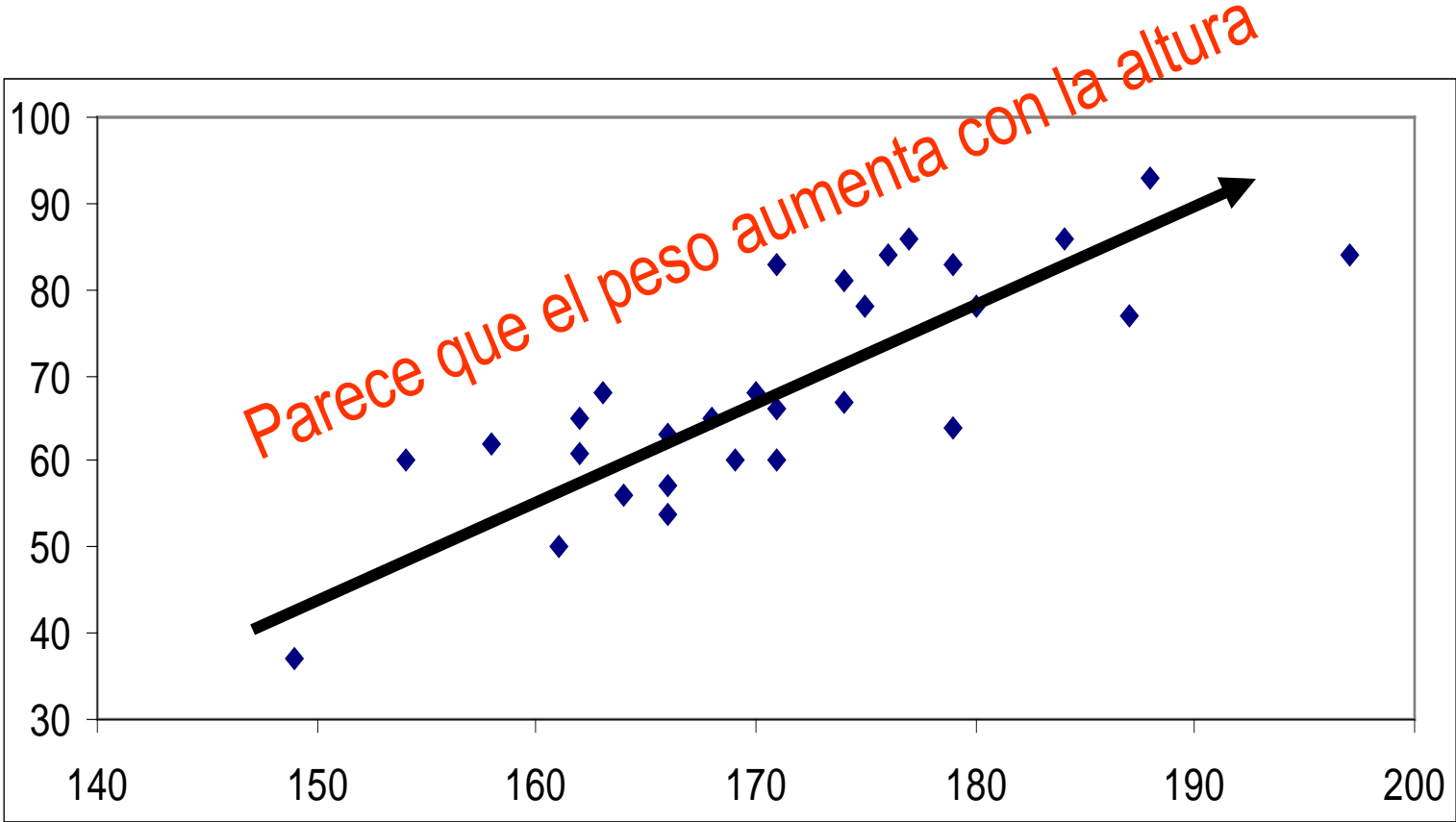
Altura en cm.	Peso en Kg.
162	61
154	60
180	78
158	62
171	66
169	60
166	54
...	...

5.2.2 Asociación entre variables cuantitativas

Diagrama de dispersión: representación gráfica en la que cada punto representa a un individuo, en cuanto a su peso y su altura



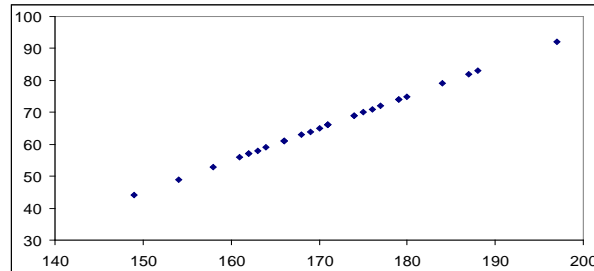
5.2.2 Asociación entre variables cuantitativas



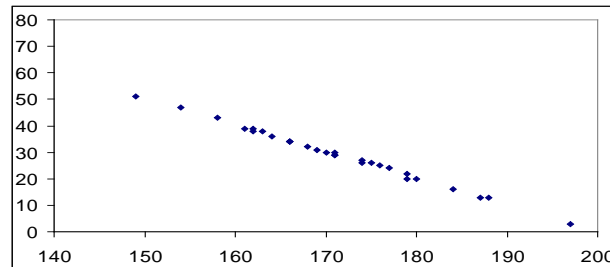
5.2.2 Asociación entre variables cuantitativas

-Relación lineal:

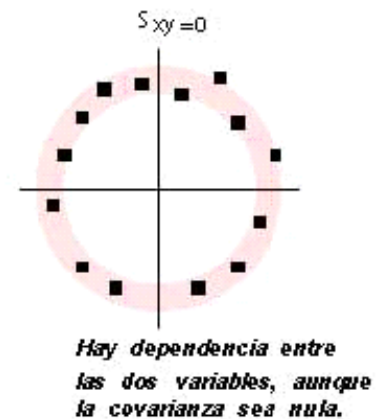
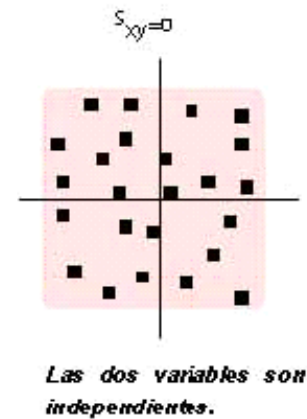
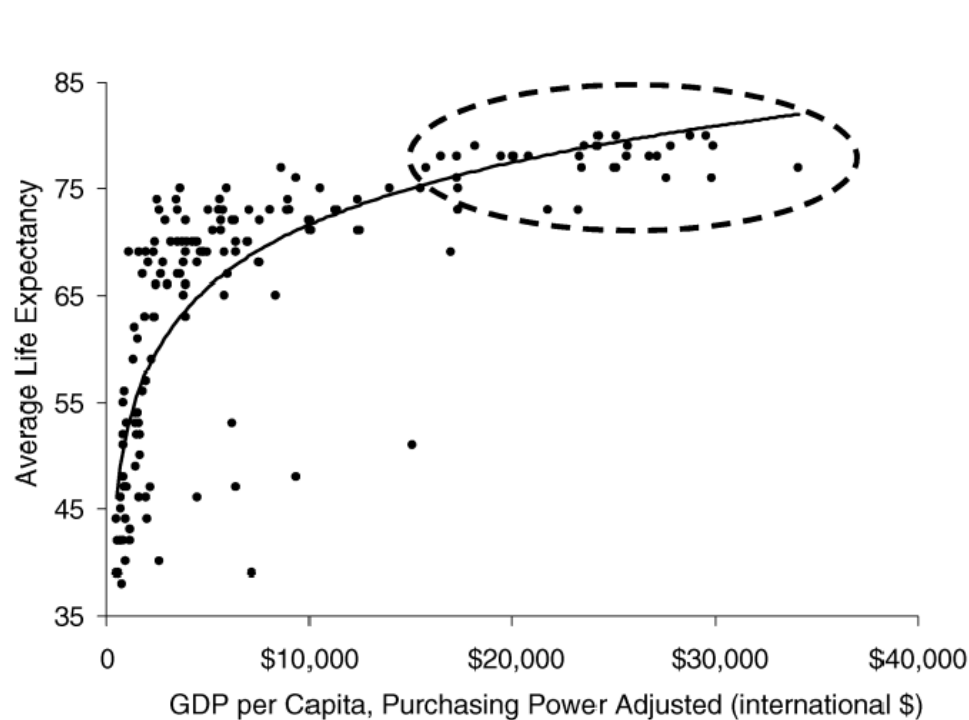
-Directa: a medida que una variable aumenta la otra también, cuando una crece la otra también



-Inversa: a medida que una variable aumenta la otra disminuye, cuando una crece la otra decrece



5.2.2 Asociación entre variables cuantitativas



También puede existir asociación entre variables cualitativas pero que no sea lineal

5.2.2 Asociación entre variables cuantitativas

- Coeficientes de correlación: Coeficiente de correlación de Pearson

Pearson investigó si existía relación entre la altura del hijo/a y la del padre, investigando más de 1000 casos y estableció una forma de predecir la altura del hijo/a en base a la del padre (regresión)



- El coeficiente de correlación de Pearson mide la relación lineal entre variables cuantitativas. Está basado en la covarianza.

5.2.2 Asociación entre variables cuantitativas

Covarianza:
$$S_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

Valor	Significado
0	No hay relación
Positivo >0	Relación directa
Negativo <0	Relación inversa

El signo de la covarianza nos dice el sentido de la relación, pero no el grado de esa relación

Covarianza:

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

Altura	Diferencia media
162	-3,7
154	-11,7
180	14,3
158	-7,7
171	5,3
169	3,3
166	0,3
Media	165,7

Peso	Diferencia media
61	-2
60	-3
78	15
62	-1
66	3
60	-3
54	-9
Media	63

Producto diferencias
7,4
35,1
214,3
7,7
15,9
-9,9
-2,6
Total=268

$$S_{xy} = \frac{268}{7} = 38,29$$

5.2.2 Asociación entre variables cuantitativas

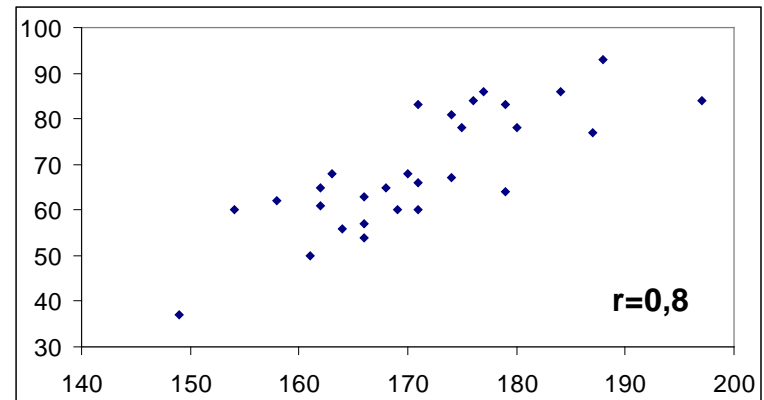
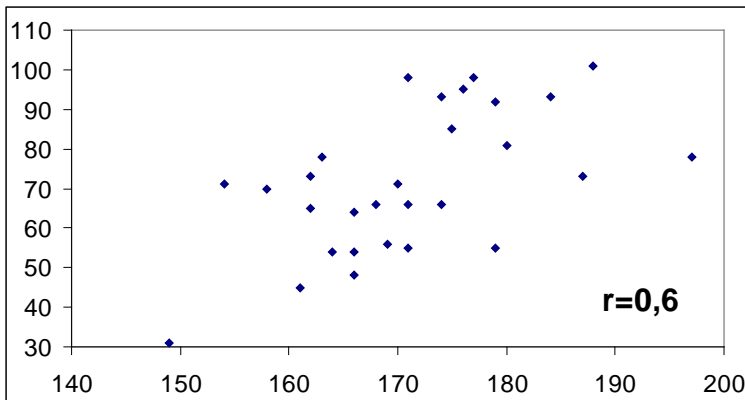
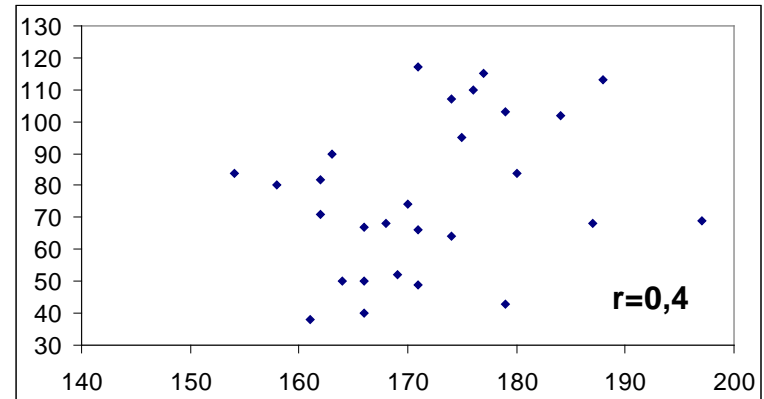
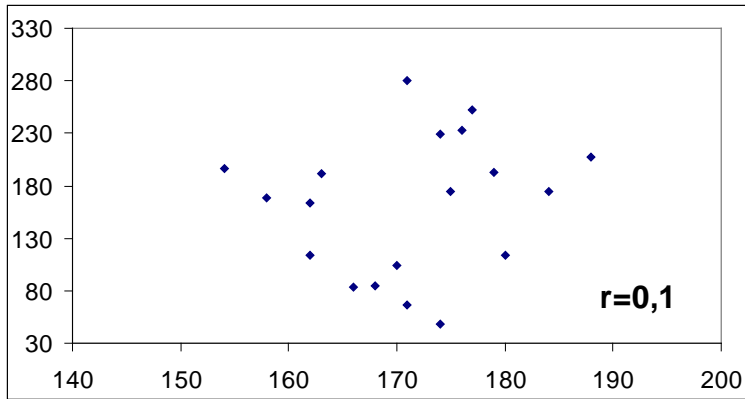
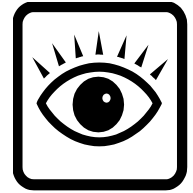
Coeficiente de correlación lineal de Pearson:

- Determina si existe relación lineal entre dos variables.
- Toma valores entre -1 y 1 (adimensional)
 - -1 relación lineal inversa
 - 0 no existe relación
 - 1 relación lineal directa

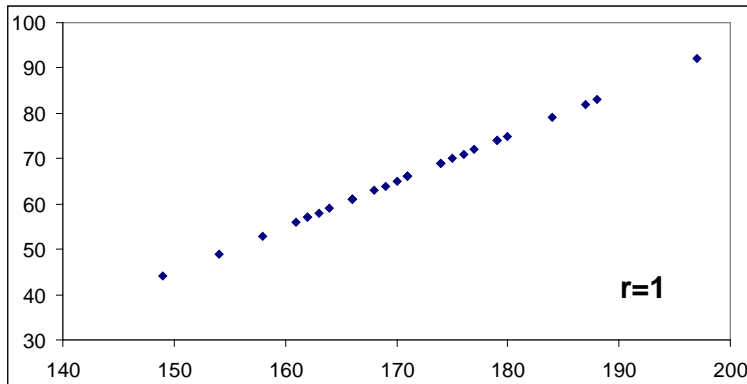
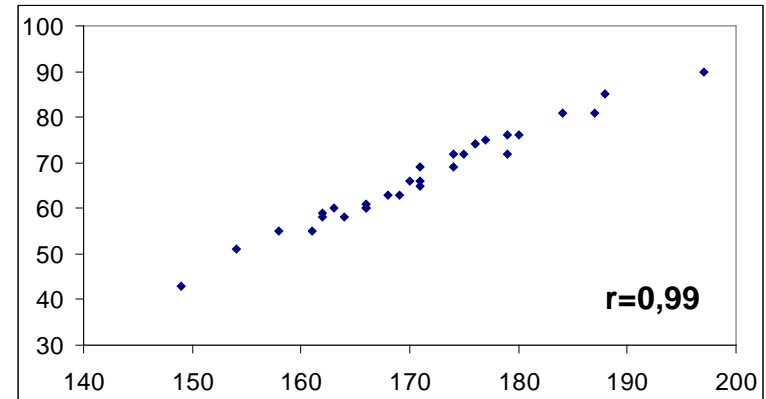
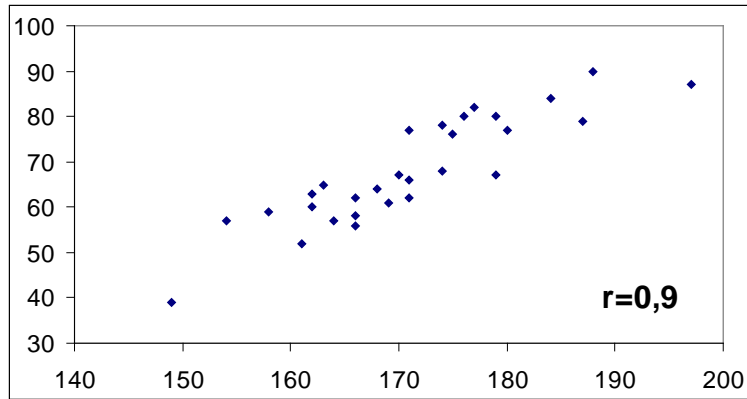
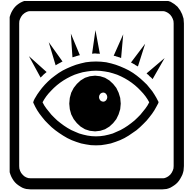
$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

- Válido para variables que siguen una distribución normal
- Existen otros coeficientes

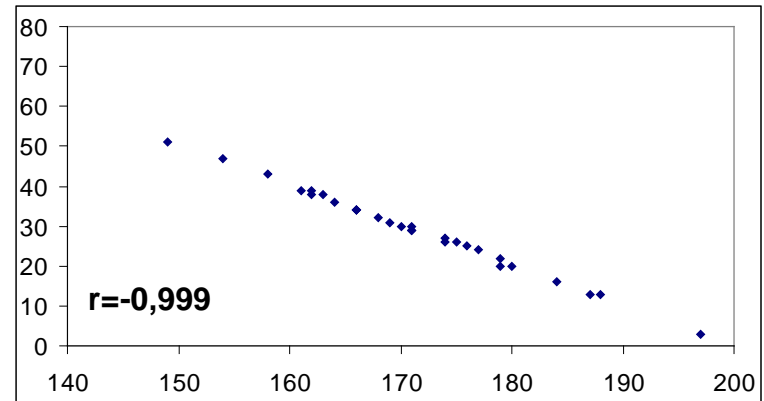
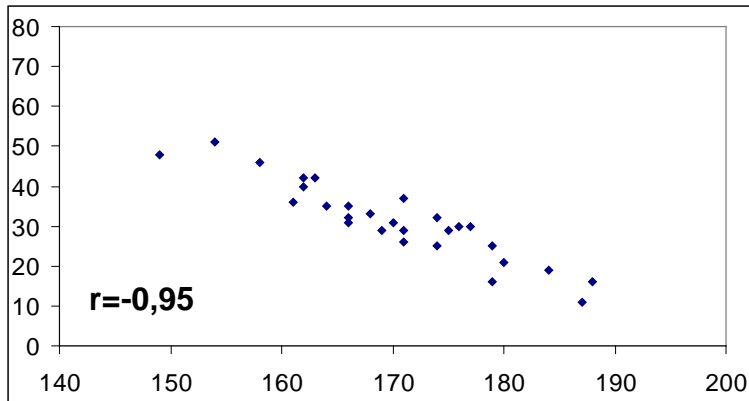
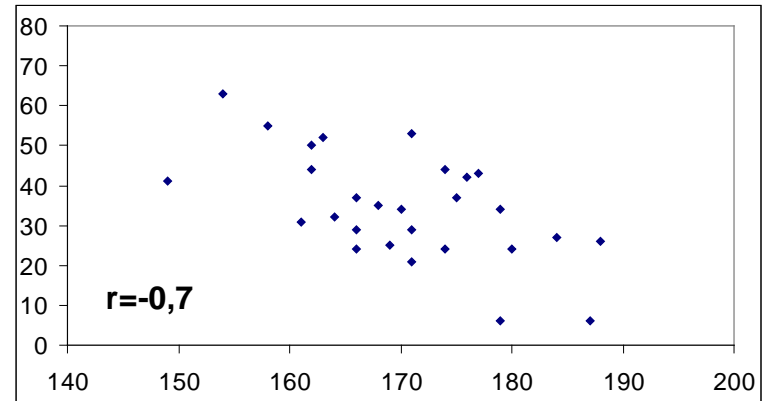
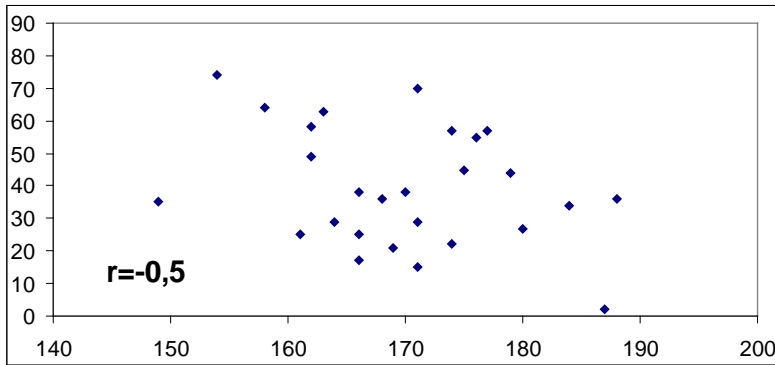
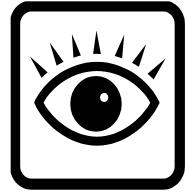
Algunas correlaciones



Algunas correlaciones



Algunas correlaciones



5.2.2 Asociación entre variables cuantitativas

Coeficiente de determinación: porcentaje de variabilidad de una variable que es explicada por la otra.

$$R^2 = r_{xy}^2$$

X= Altura Y= peso

$r_{xy} = 0,95$ Existe una correlación lineal fuerte entre la altura y el peso, en el sentido de que a más altura más peso.

$R^2 = 0,90$ El 90% de la variación en el peso se debe a la variación en la altura

5.2.3 Asociación entre variables cualitativas

A la hora de estudiar la asociación entre dos variables cualitativas contamos con:

- Tablas de contingencia
- Coeficiente de contingencia

5.2.3 Asociación entre variables cualitativas

¿Depende la elección de grado del sexo?

Matriculación según grado y sexo

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	3	5	64	72
Mujer	45	72	12	129
Total	48	77	76	201

5.2.3 Asociación entre variables cualitativas

Para que hablemos de **dependencia** entre dos variables, tiene que darse:

1º Los porcentajes por columna sean diferentes

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	6	6	84	36
Mujer	94	94	16	64
Total	100	100	100	100

Dependencia: el sexo influye en la elección de la carrera

¿Qué significa el 16?: El 16% del alumnado de filosofía son mujeres

5.2.3 Asociación entre variables cualitativas

2º Los porcentajes por fila sean diferentes

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	4	7	89	100
Mujer	35	56	9	100
Total	24	38	38	100

Dependencia: el sexo influye en la elección de la carrera

¿Qué significa el 9?: El 9% de las mujeres estudian filosofía

5.2.3 Asociación entre variables cualitativas

3º Los valores reales y esperados sean diferentes

(valor esperado en el caso de independencia absoluta)

$$n_{11} = (n_{1\cdot} * n_{\cdot 1}) / n$$

$$n_{11} = (48 * 72) / 201$$

Valores reales

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	3	5	64	72
Mujer	45	72	12	129
Total	48	77	76	201

≠

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	17,19	27,58	27,22	72
Mujer	30,81	49,42	48,78	129
Total	48	77	76	201

Dependencia: el sexo influye en la elección de la carrera

5.2.3 Asociación entre variables cualitativas

Coeficiente de contingencia (C): nos mide la asociación o dependencia entre las dos variables.

Ofrece valores entre 0 y 1, siendo 0 independencia total y 1 dependencia total

Es un coeficiente conservador

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \longrightarrow \chi^2 = \sum \frac{(\text{frec.observada} - \text{frec.esperada})^2}{\text{frc.esperada}}$$

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$\chi^2 = \sum \frac{(\text{frec.observada} - \text{frec.esperada})^2}{\text{frc.esperada}}$$

Tabla real

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	3	5	64	72
Mujer	45	72	12	129
Total	48	77	76	201

Tabla valores esperados

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	17,19	27,58	27,22	72
Mujer	30,81	49,42	48,78	129
Total	48	77	76	201

$$\chi^2 = \frac{(3 - 17,19)^2}{17,19} + \frac{(5 - 27,58)^2}{27,58} + \dots + \frac{(12 - 48,78)^2}{48,78}$$

	Pedagogía	Ed. Social	Filosofía	Total
Hombre	11,7	18,5	49,7	
Mujer	6,5	10,3	27,7	
Total				124,5

$$C = \sqrt{\frac{124,5}{201 + 124,5}}$$

Por último ...

- Cuidado con terceras variables que afectan la relación entre las dos variables que estudiamos

- Correlación no siempre significa causa

 - Dos variables varían de forma simultánea

 - Consecuencia de terceras variables

 - Para que una sea causa y la otra efecto:

 - Debe haber relación

 - Una debe suceder antes que la otra

 - No hay terceras variables

Por último...

