Bioestadística

Práctica de Ordenadores 3

José Aurelio Pina Romero
Ja.pina@ua.es
Bioestadística – Grado Enfermería
UA- Departamento de Enfermería

Universitat d'Alacant
Universidad de Alicante

¿Qué vamos hacer?

- Ejercicio 1.19 (15 minutos) Repaso P02
- Ejercicio 1.7 - Profe
- Ejercicio 1.15 (+ regresión lineal) - Profe
- Ejercicio 1.23 (+regresión lineal) - Clase
- Opcional/casa: ejercicio 1.6

Ejercicio 1.19

A partir de la base de datos LITIABIS.sav

i. calcule los estadísticos descriptivos de tendencia central, dispersión y forma para las variables EDAD, TALLA, PESO y TRIGLICERIDOS. Interpreta los resultados.
ii. Calcule los percentiles 5,25, 75 y 95. Interpreta de los resultados.

Ejercicio 1.7

Se determinó el peso (en Kg) y el nivel de glucosa en la sangre (en mg/100 ml) de 18 hombres y 18 mujeres que trabajaban en una fábrica de bombones. Los resultados obtenidos fueron los siguientes:

Ejercicio 1.7

HOMBRES

PESO 75.0 62.5 80.0 76.0 68.5 60.0 61.5 68.5 70.0 72.0 74.0 76.5 65.5 64.5 67.0 82.0 75.0 72.5

GLUCOSA 90 87 108 95 89 85 95 100 105 100 104 107 98 100 105 110 102 99

MUJERES

PESO 55.0 49.5 70.0 60.0 58.5 50.0 51.5 58.5 59.0 62.0 52.0 66.5 55.5 57.5 47.0 72.0 65.0 61.5

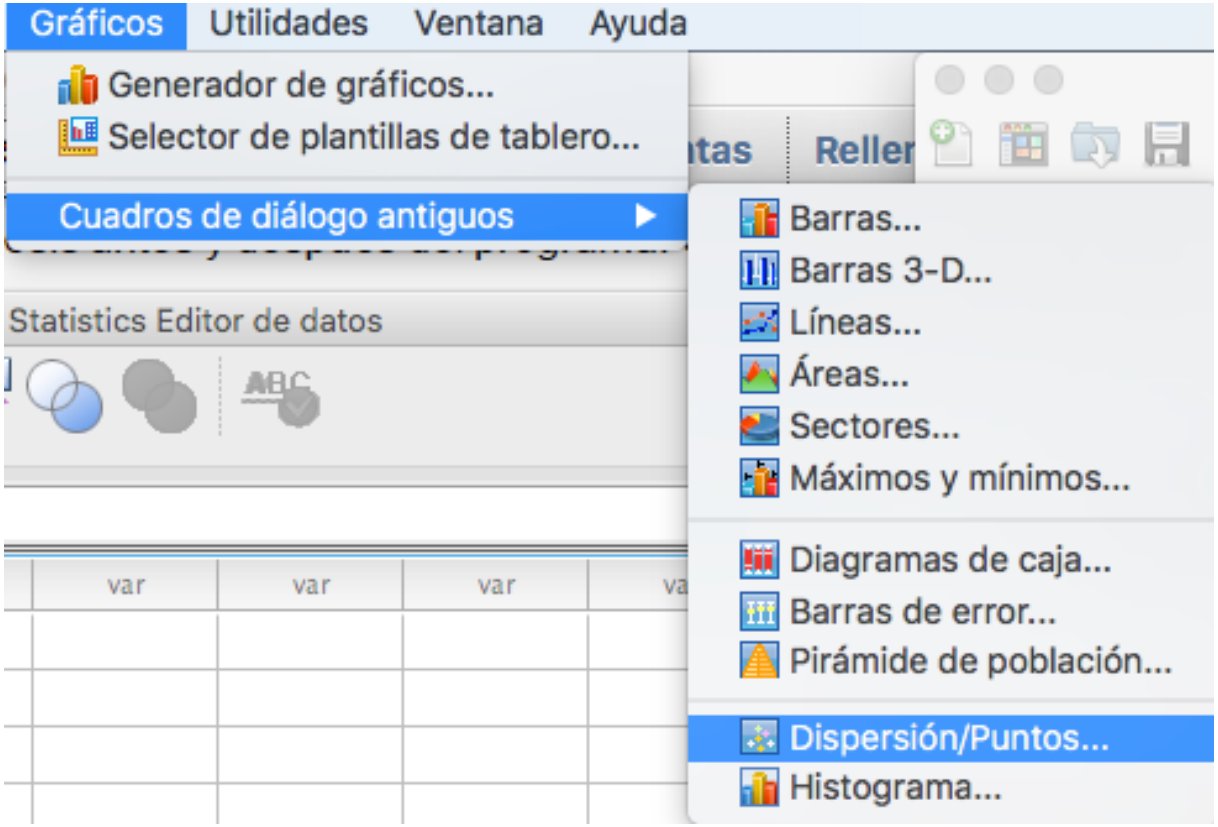
GLUCOSA 100 97 118 105 99 95 102 110 115 103 108 117 98 105 109 115 102 107

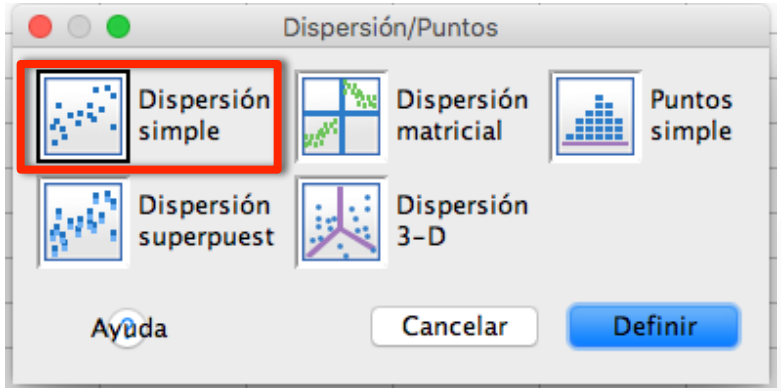
Ejercicio 1.7

Construya el diagrama de dispersión para las variables Peso y Glucosa tanto para hombres como para mujeres. (Se sugiere representar ambos sexos sobre el mismo diagrama de dispersión distinguiendo mediante distintos colores).

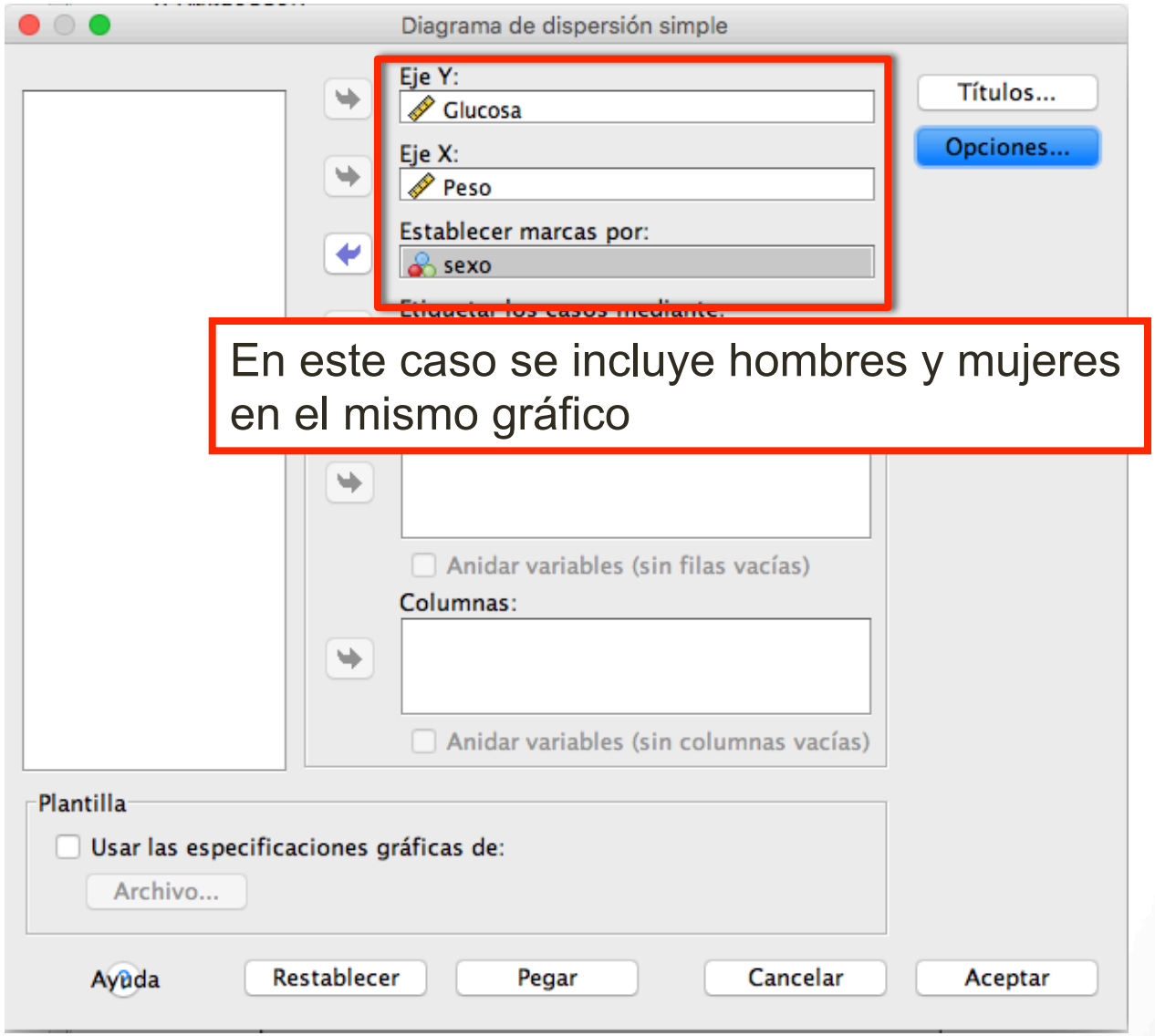
Interprete los resultados.

Diagramas de dispersión con SPSS



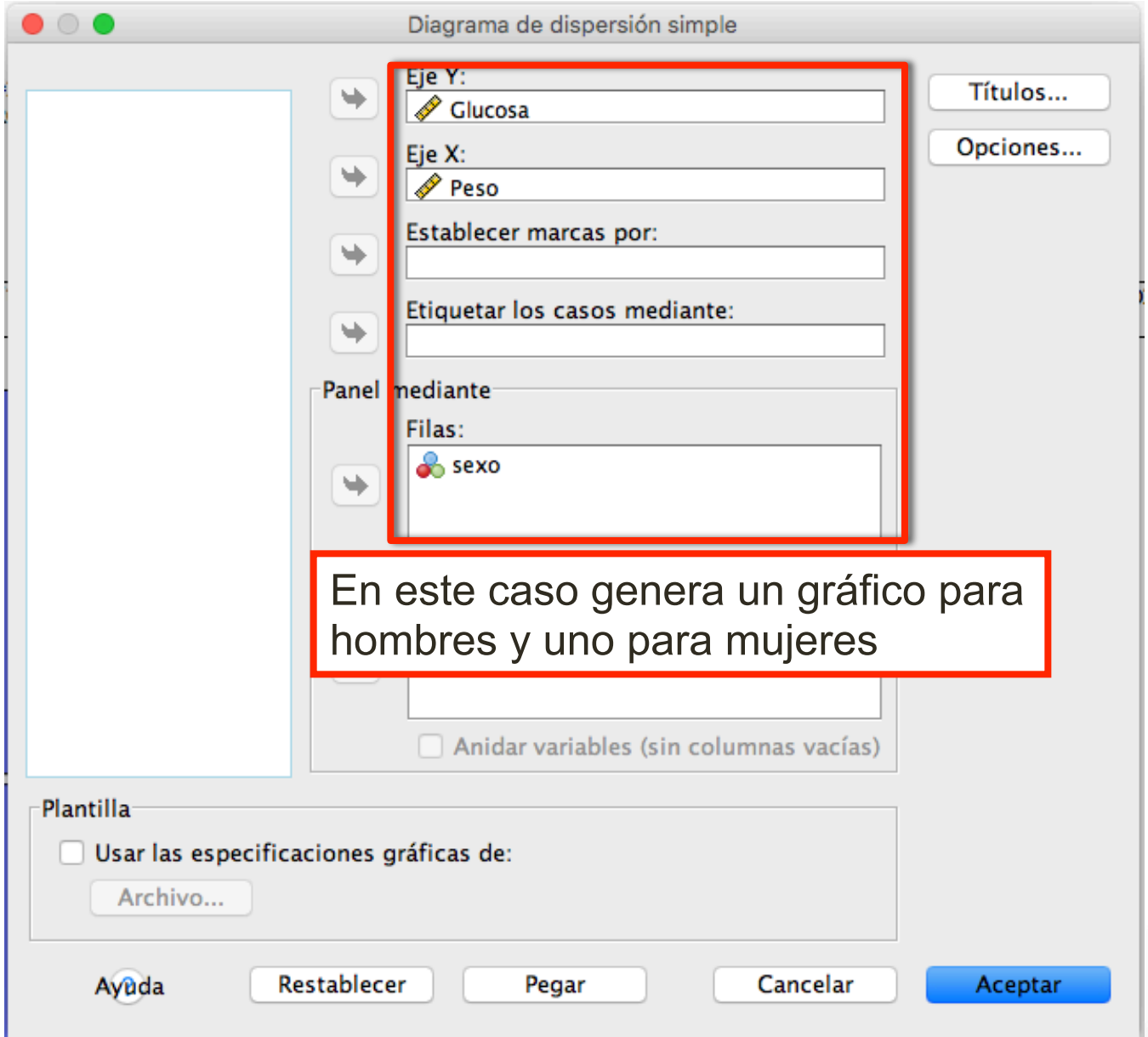


Diagramas de dispersión con SPSS



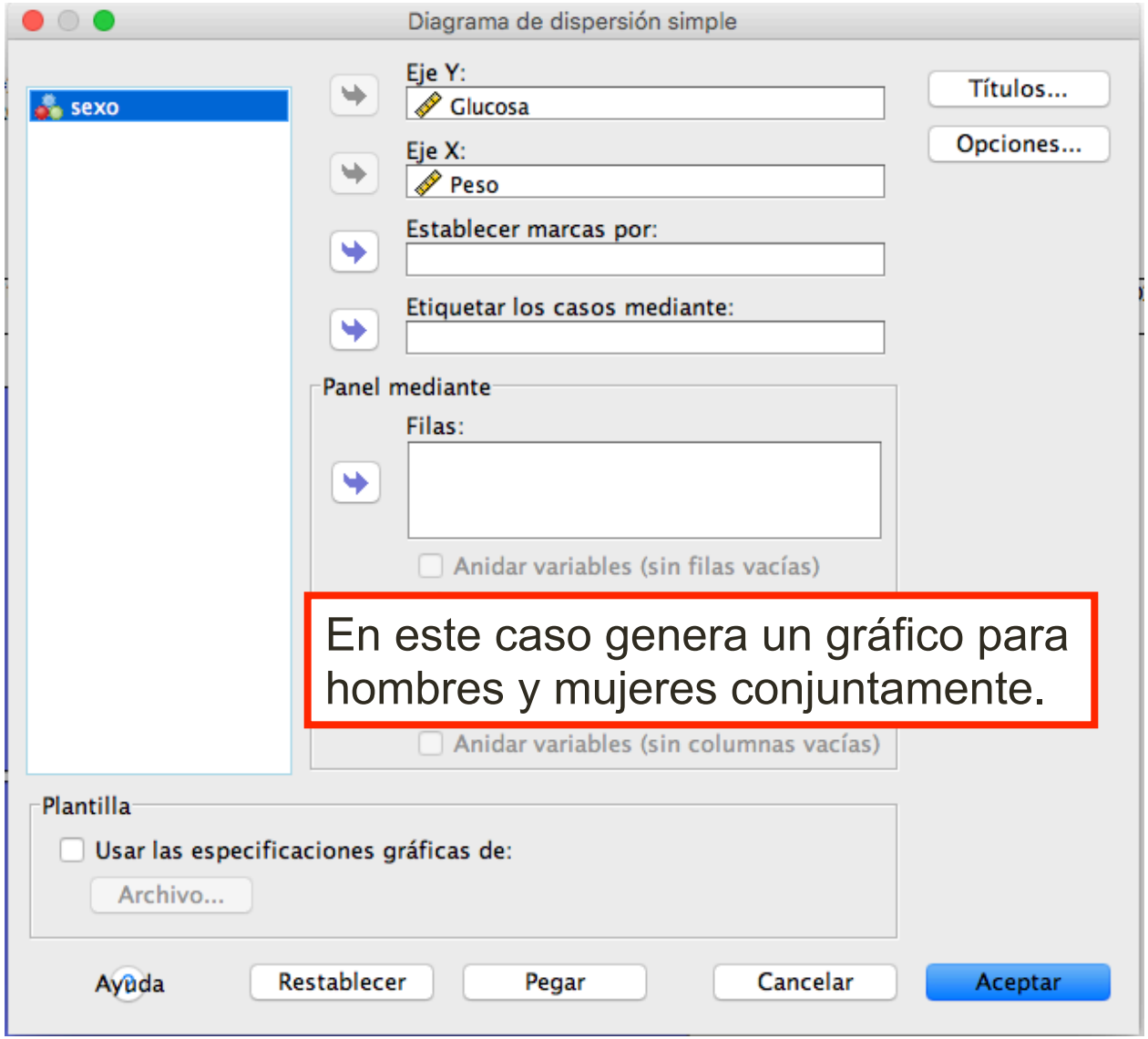
En este caso se incluye hombres y mujeres en el mismo gráfico

Diagramas de dispersión con SPSS



En este caso genera un gráfico para hombres y uno para mujeres

Diagramas de dispersión con SPSS



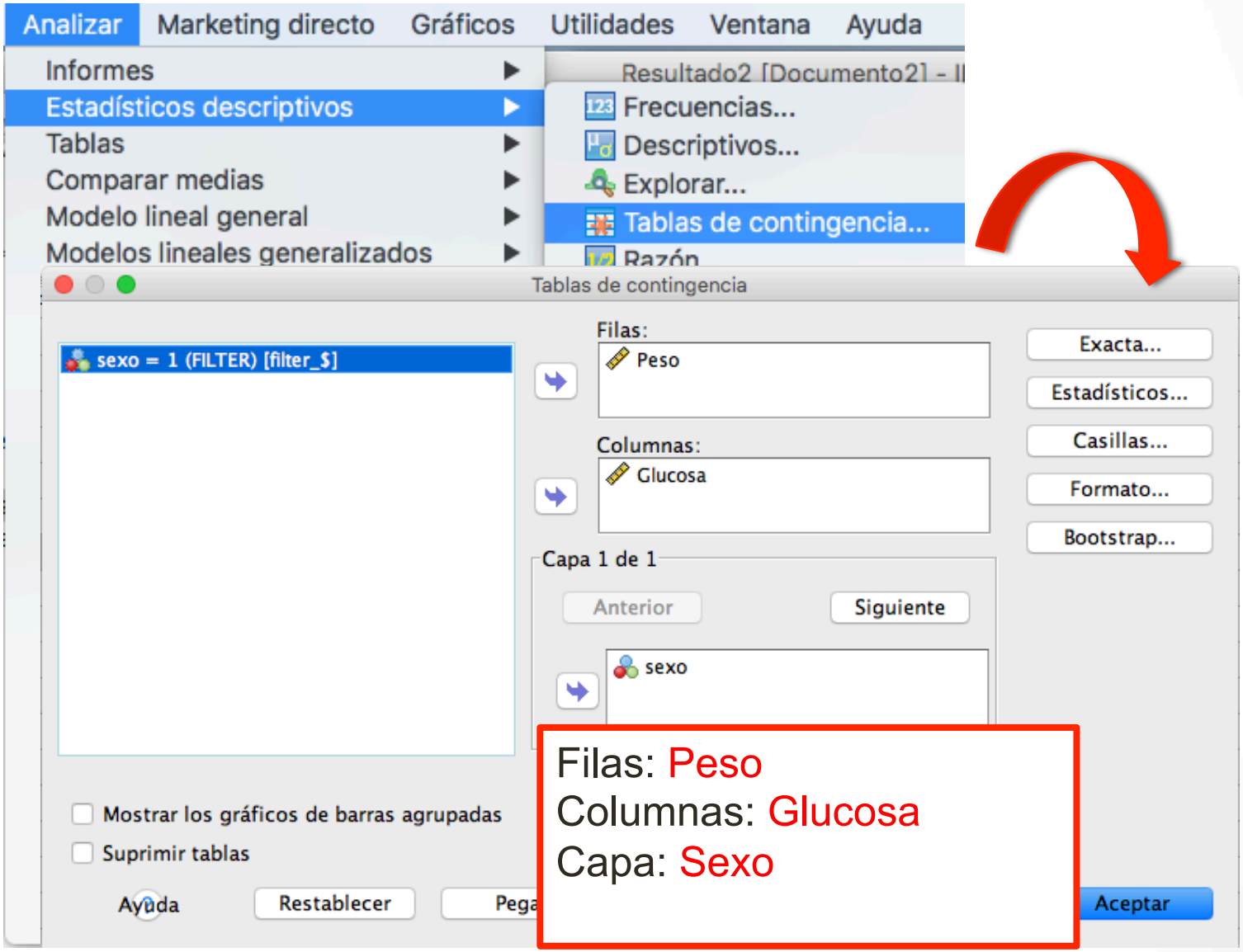
En este caso genera un gráfico para hombres y mujeres conjuntamente.

Ejercicio 1.15

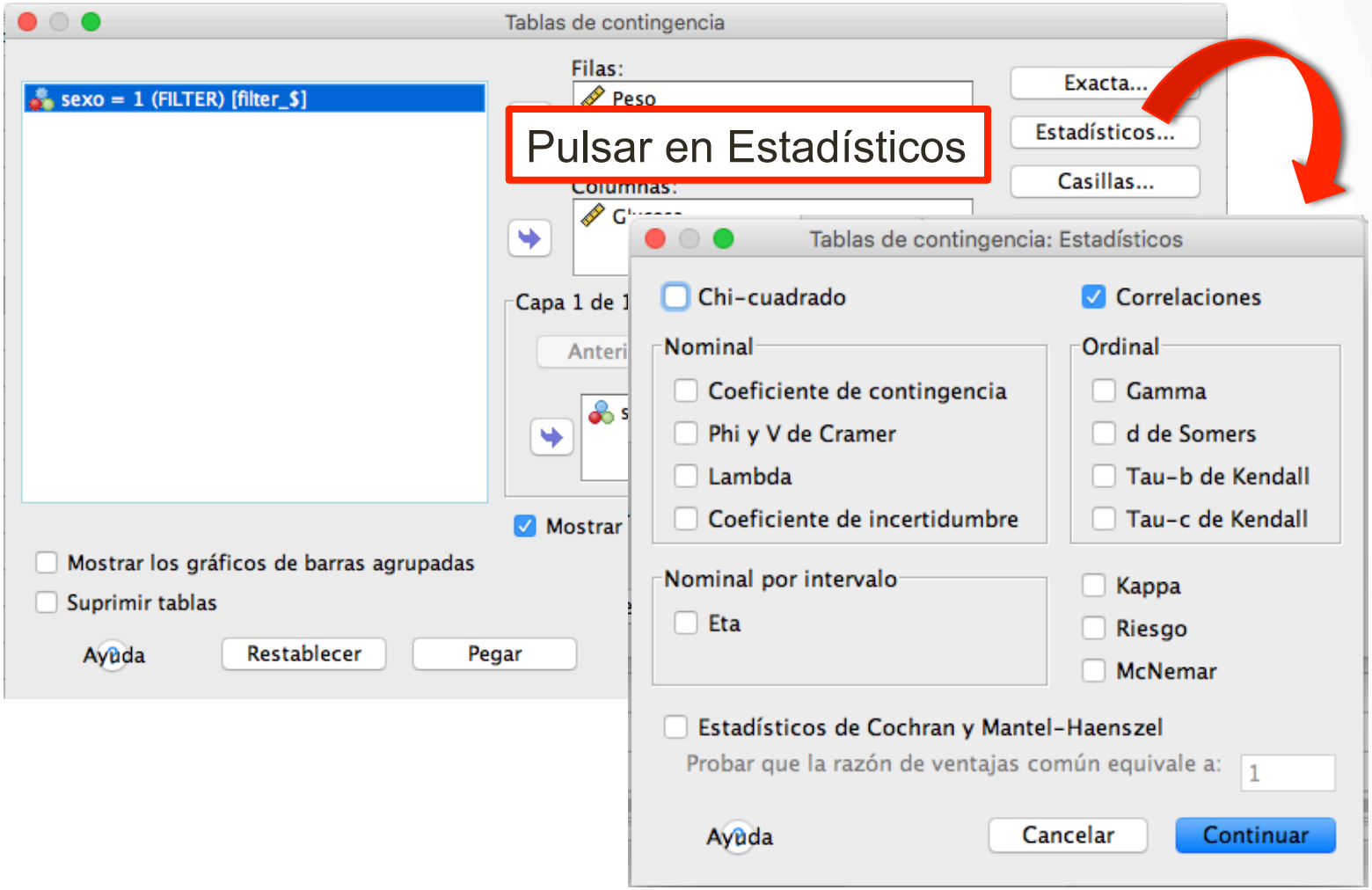
Basándose en los datos del ejercicio 1.7, calcule la **correlación lineal** entre las variables peso y glucosa.

- i. Para hombres y mujeres conjuntamente
- ii. Para hombres y mujeres por separado
- iii. Compare los resultados
- iv. Construya un modelo de regresión lineal

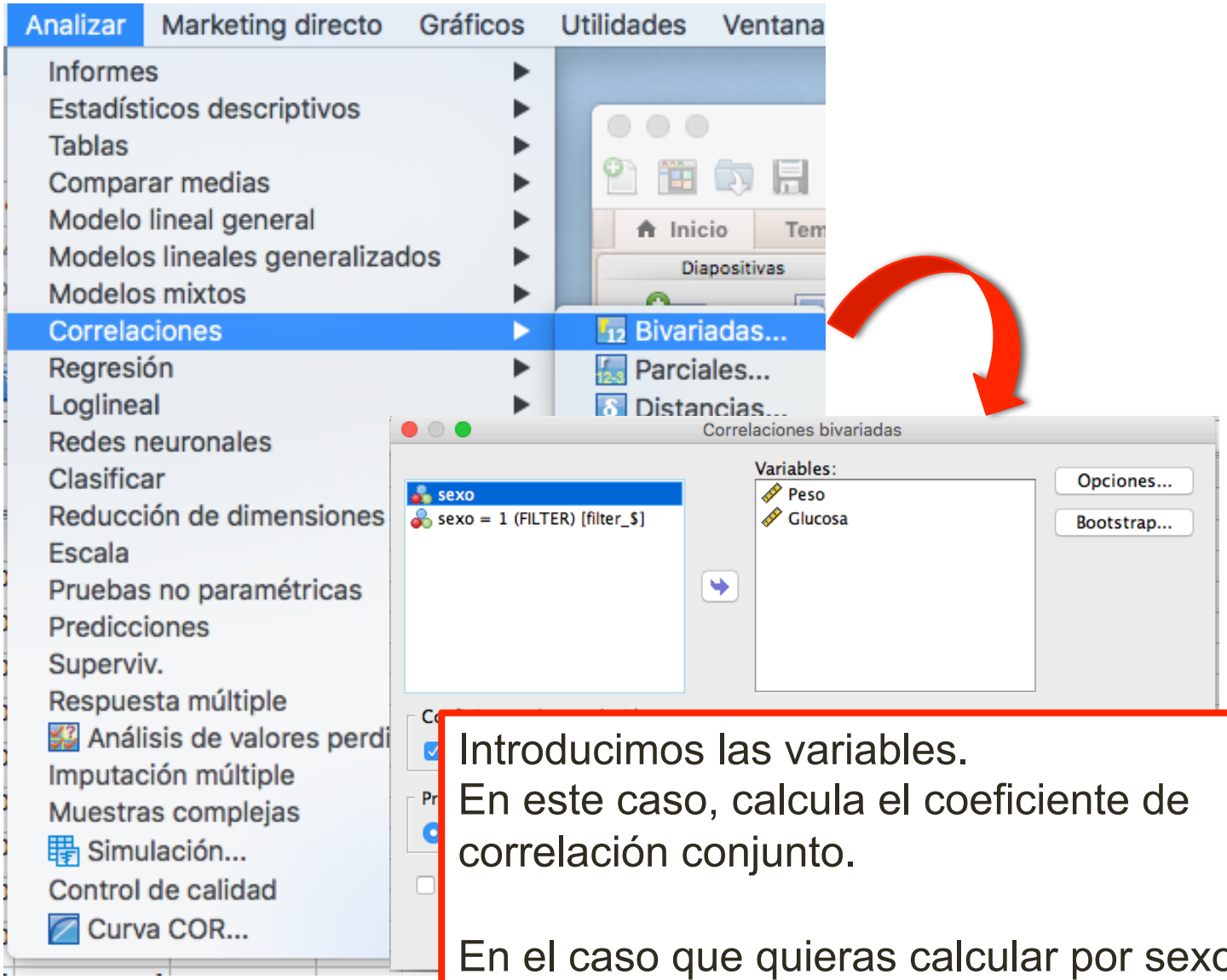
Correlación lineal SPSS



Correlación lineal SPSS



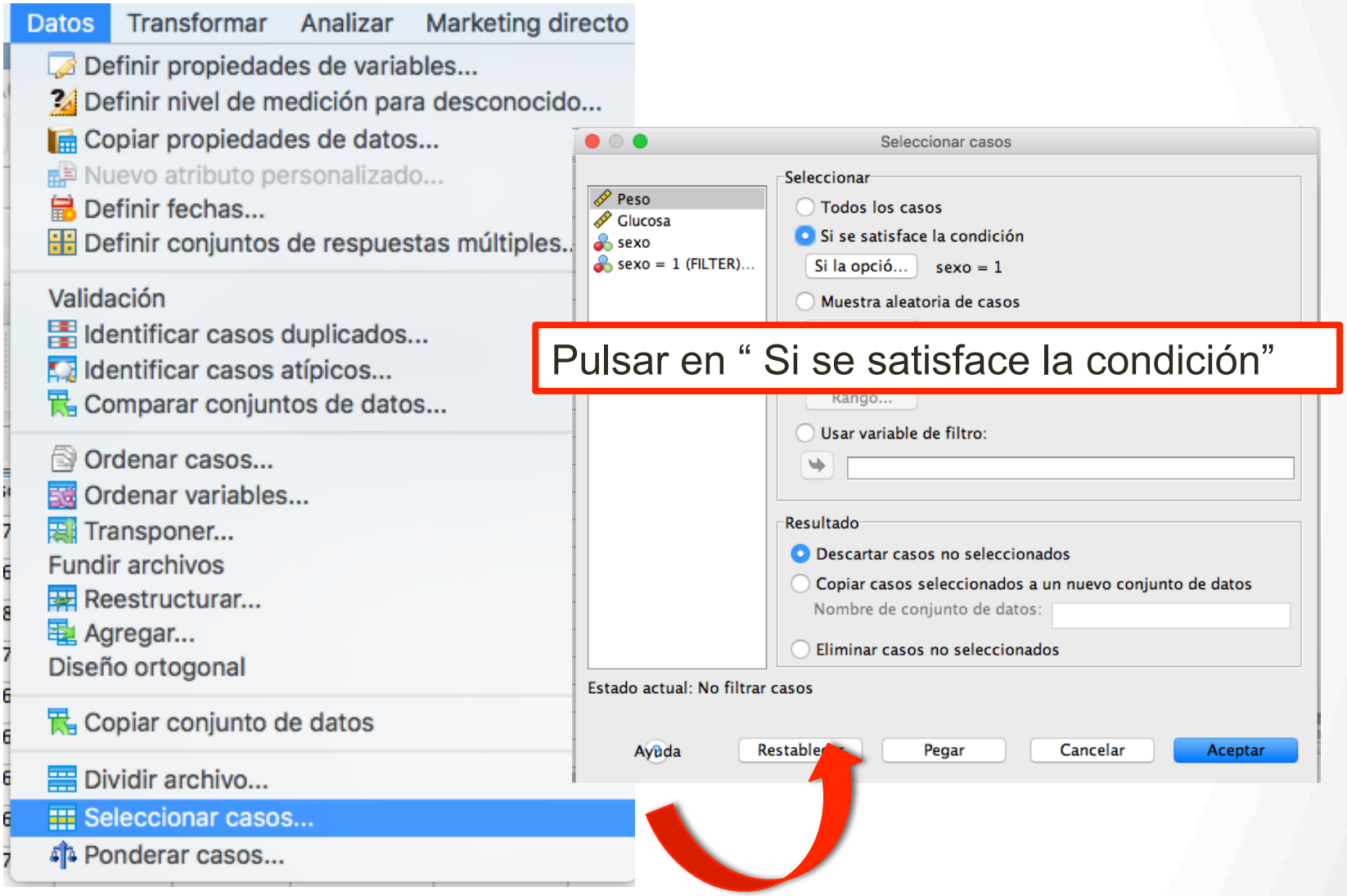
Correlación lineal SPSS – Otra forma



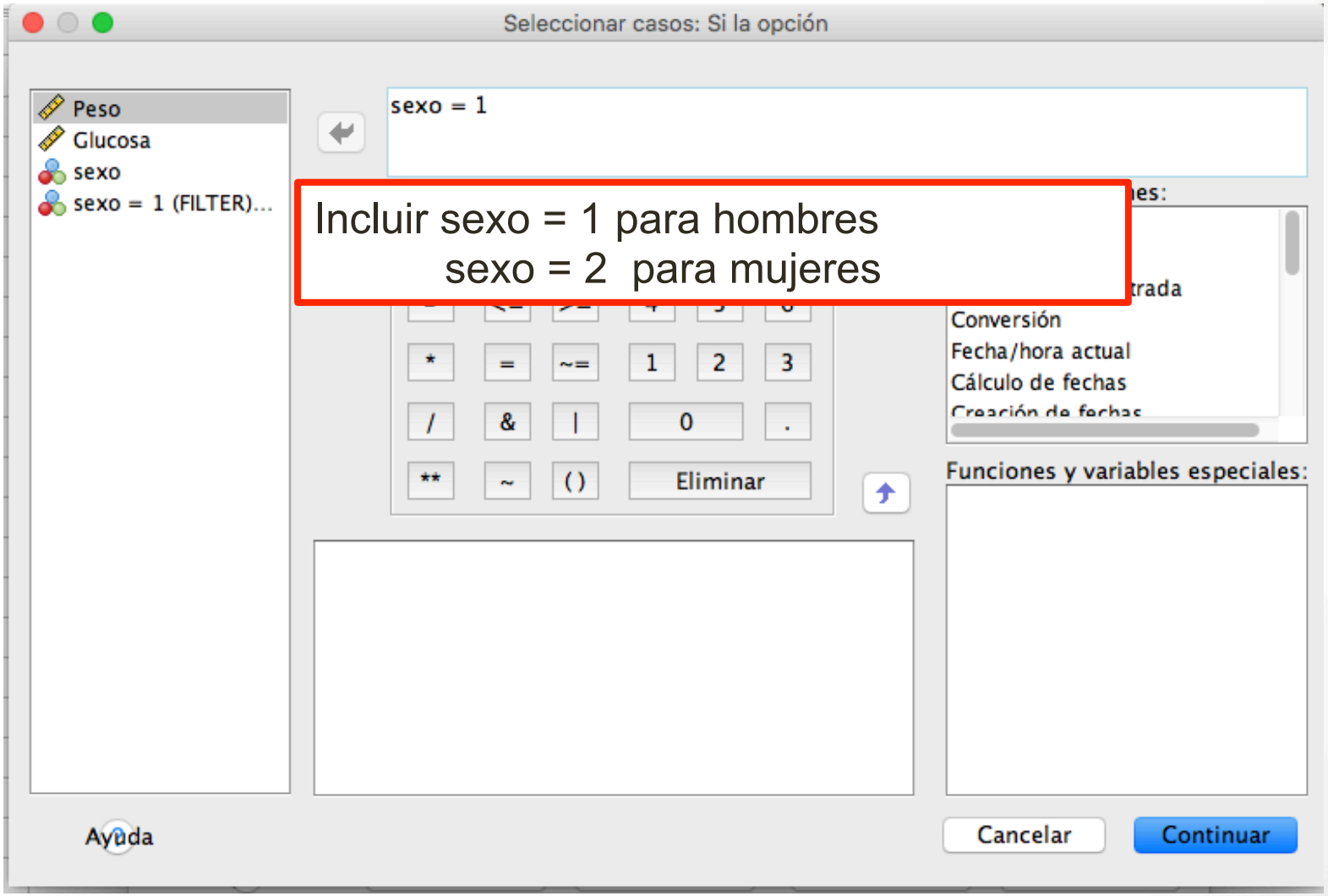
Introducimos las variables.
En este caso, calcula el coeficiente de correlación conjunto.

En el caso que quieras calcular por sexo, hay que seleccionar los casos.

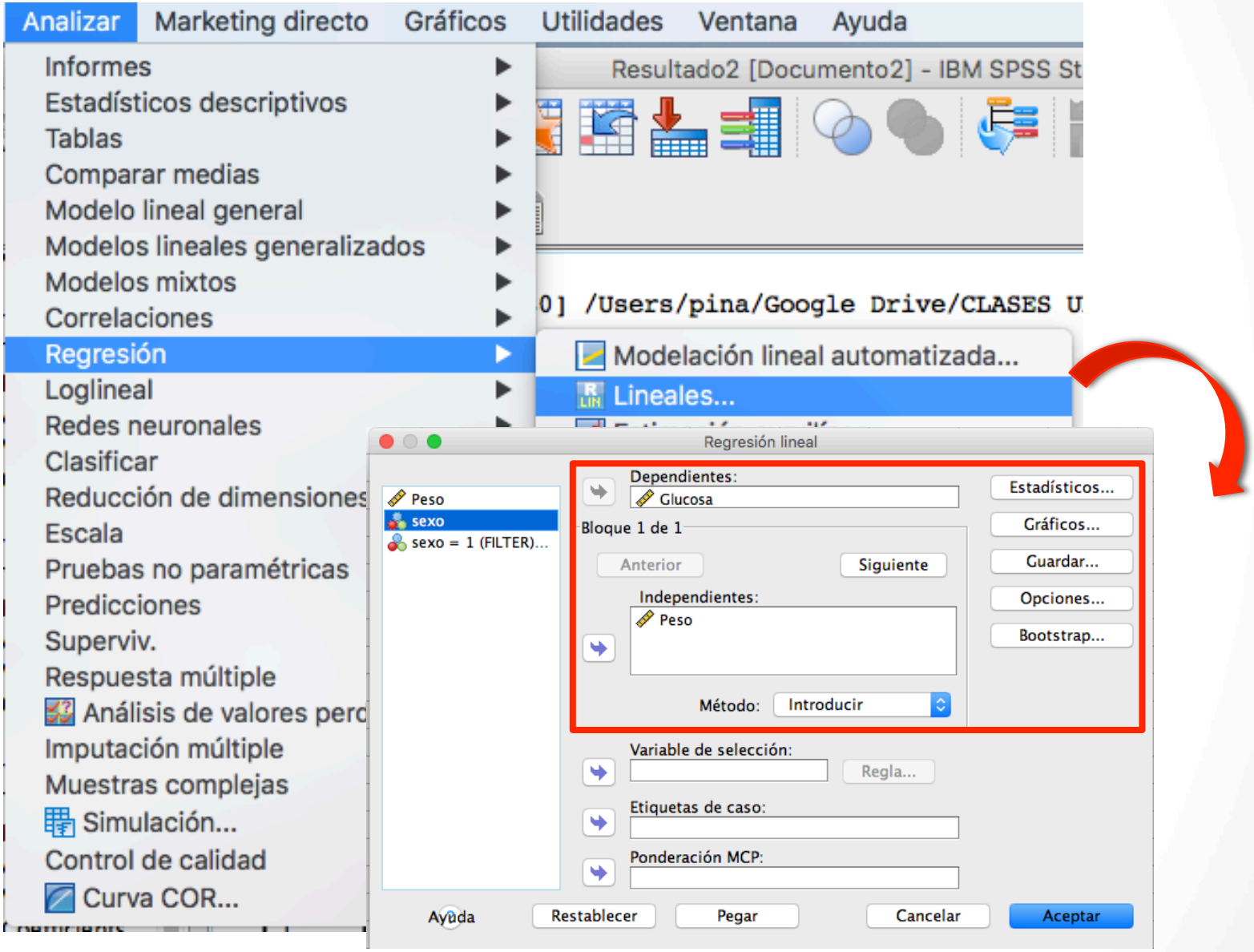
Correlación lineal SPSS – Otra forma



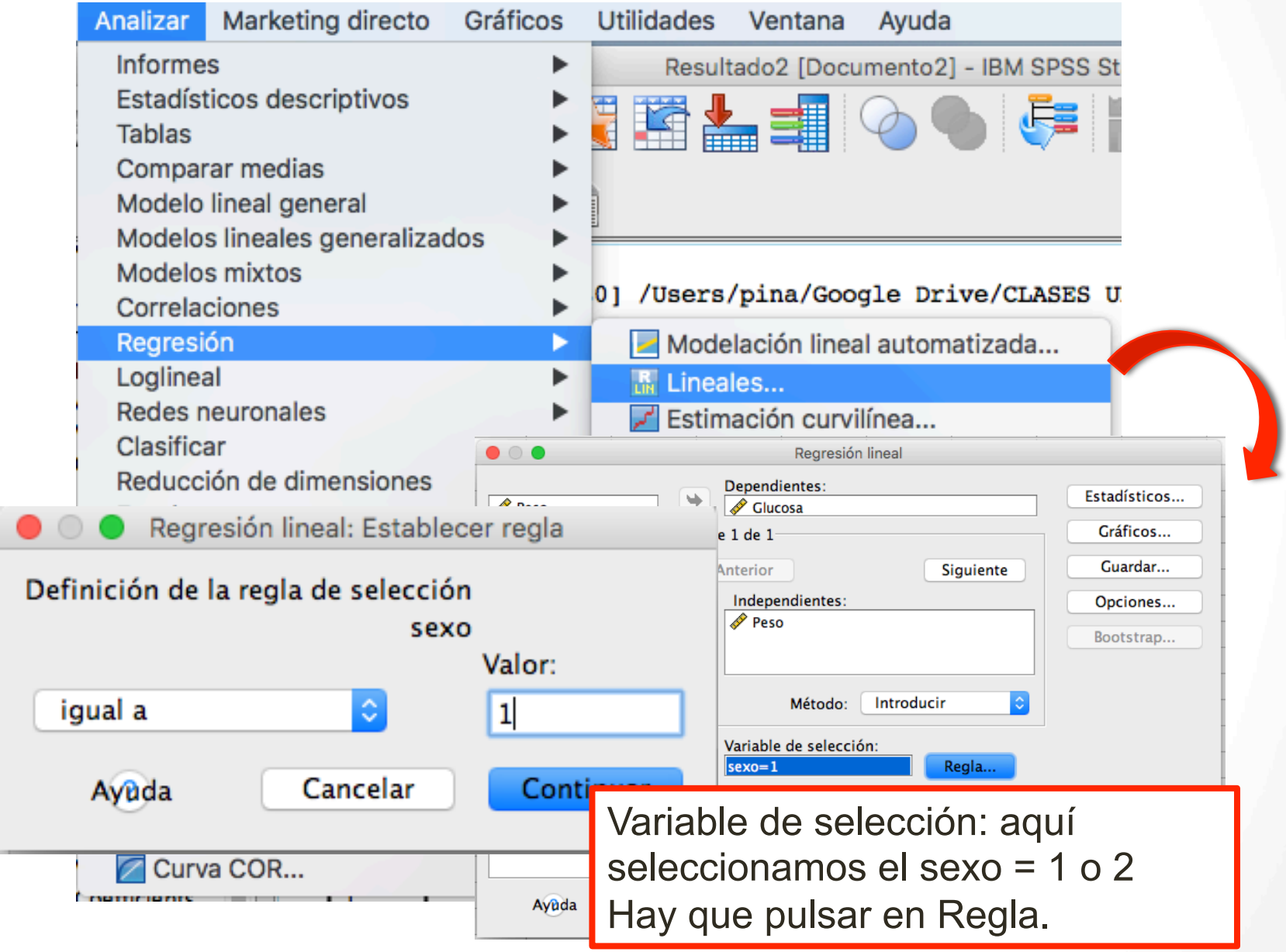
Correlación lineal SPSS – Otra forma



Modelo de regresión lineal - SPSS



Modelo de regresión lineal - SPSS



Variable de selección: aquí seleccionamos el sexo = 1 o 2
Hay que pulsar en Regla.

TEORÍA

Correlación - Regresión

Diferentes formas de describir la relación entre dos variables (estadística descriptiva bivariante) cuando estas son numéricas (cuantitativas).

- Estudiar si hay relación entre la altura y el peso
- Estudiar si hay relación entre el nivel de colesterol y la edad.
- Estudiar si hay relación entre el peso y el nivel de glucosa en sangre.

Correlación - Regresión

Objetivo:

- Intentar reconocer a partir de un diagrama de dispersión si hay relación entre las variables a estudio.
- Tipo de relación.
- Predecir, si es posible, el valor de una de ellas en función de la otra.
- Medir el grado de intensidad de esa relación.

Estudio conjunto de dos variables

Posible forma de recoger los datos.

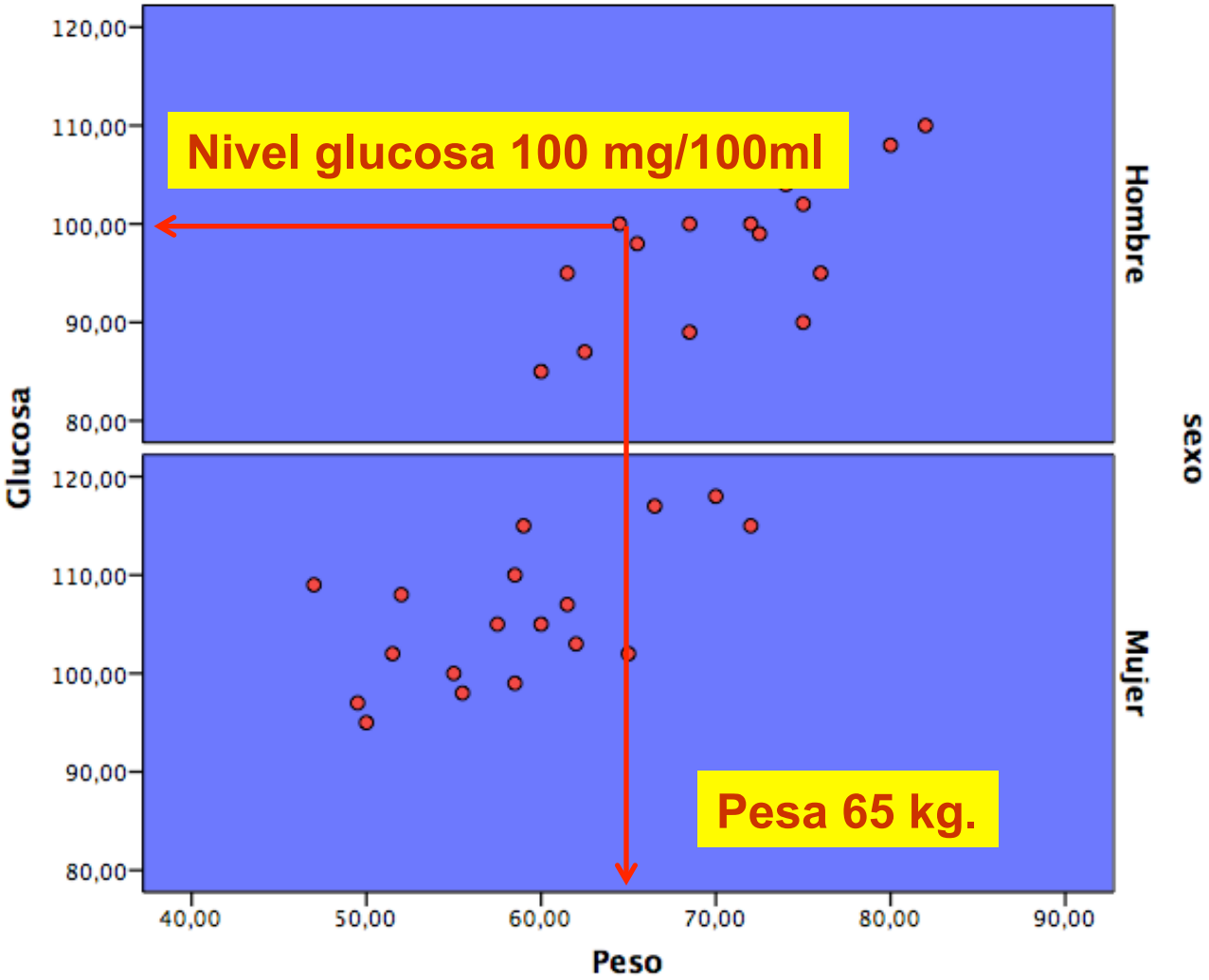
Glucosa	Peso en Kg.
87	61
108	60
95	78
89	62
85	66
95	60
100	54
105	84
100	68
...	...

En cada fila tenemos los datos de un individuo.

Y en cada columna los valores que toma cada variable.

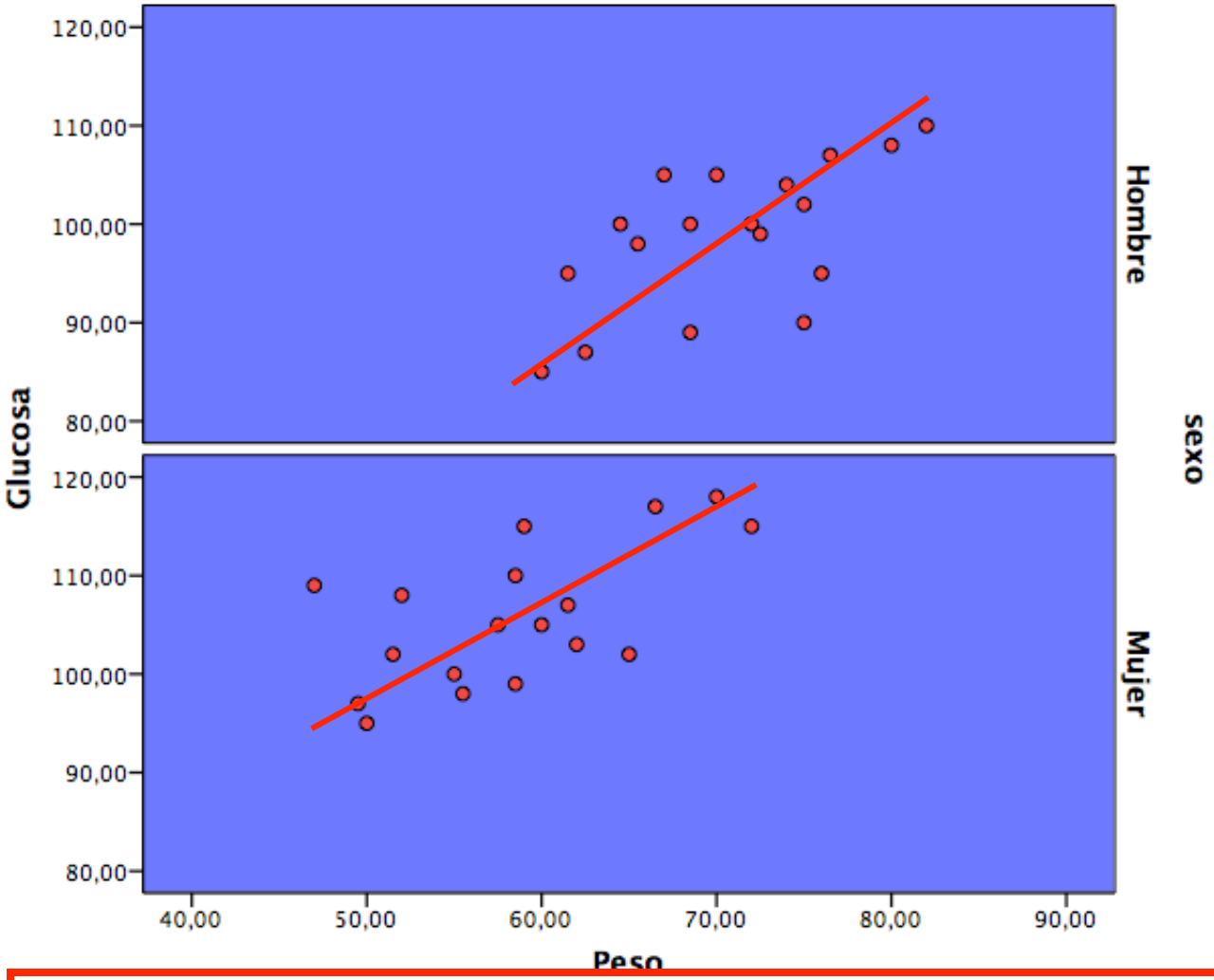
Los datos se representaran mediante un diagrama de dispersión. Cada individuo es un punto cuyas ordenadas son los valores de las variables.

Diagrama de dispersión



Tenemos el nivel de glucosa y los pesos de 18 hombres y 18 mujeres representados en un diagrama de dispersión.

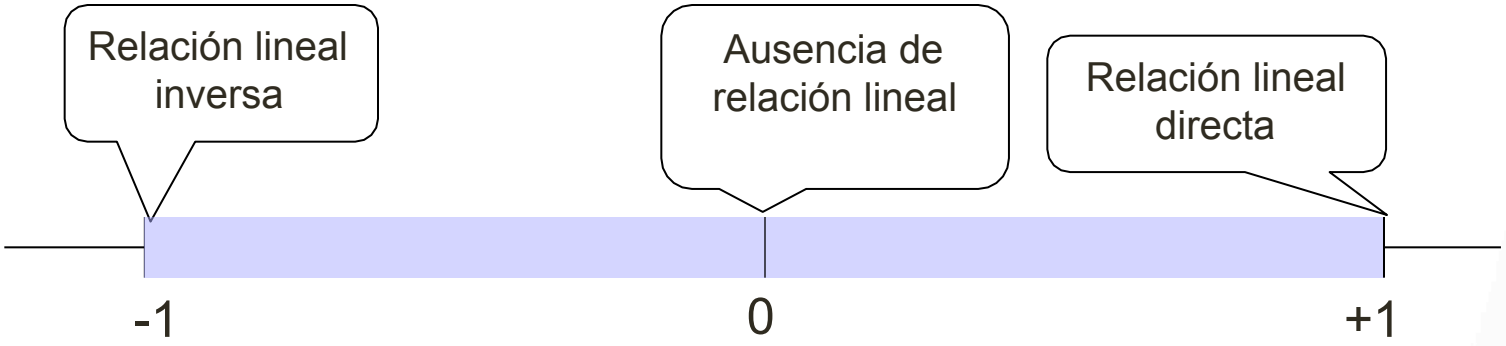
Diagrama de dispersión



Parece que el nivel de glucosa aumenta con el peso

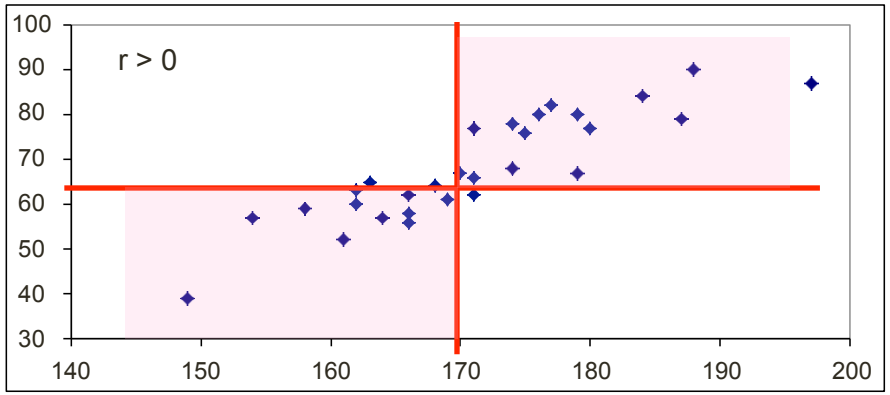
Coeficiente de correlación lineal de Pearson

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sum_{i}^{n}\left(y_i - \overline{y}\right)^2}} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

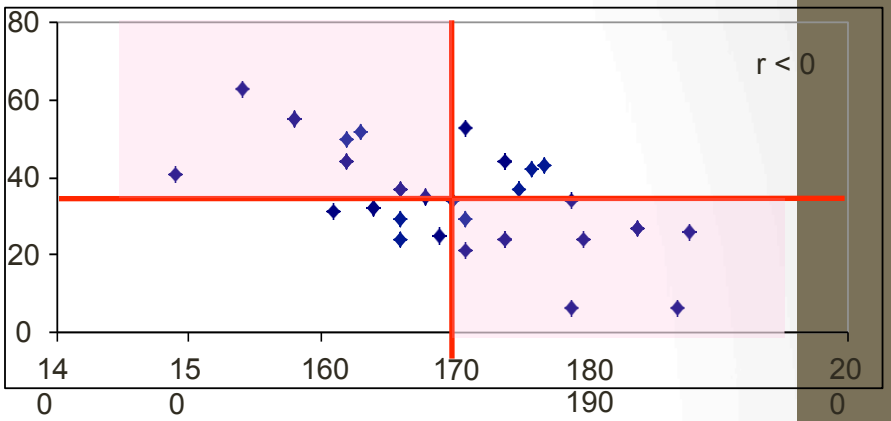


La coeficiente de correlación lineal de Pearson de dos variables nos indica si los puntos tienen una tendencia a disponerse alineadamente

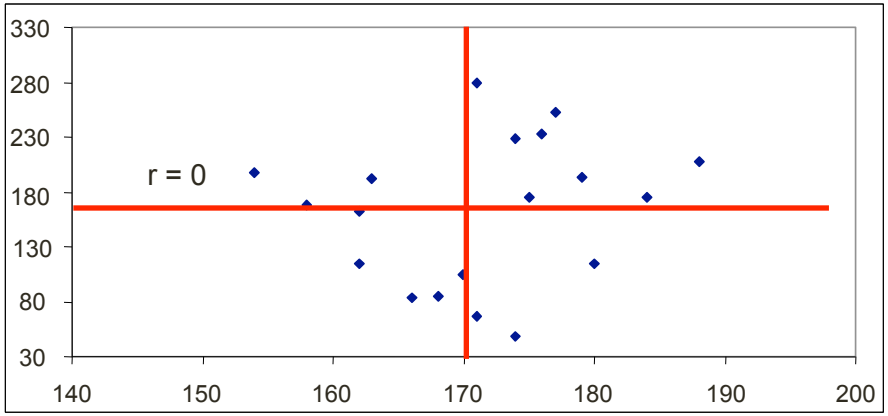
Formas de relación entre variables



Relación lineal directa.

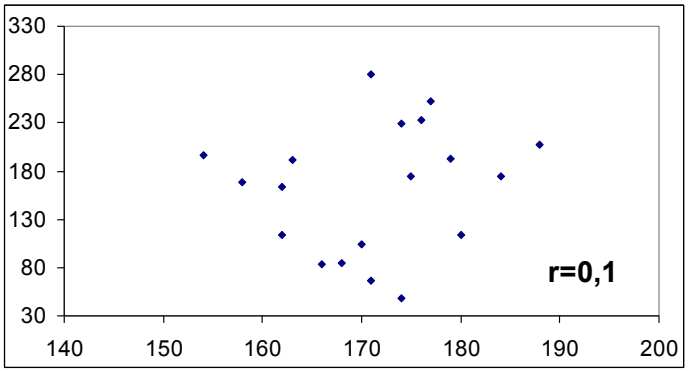


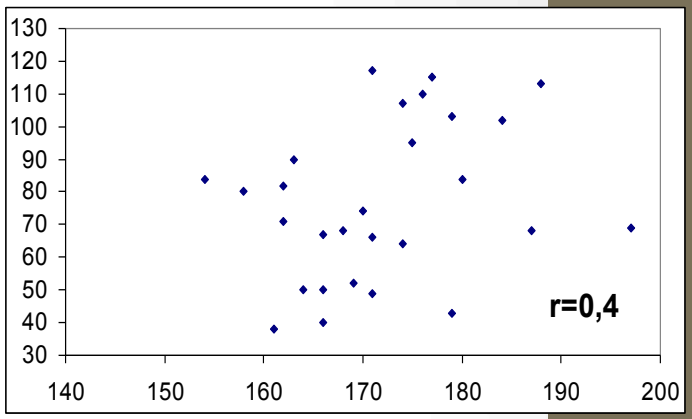
Relación lineal inversa.

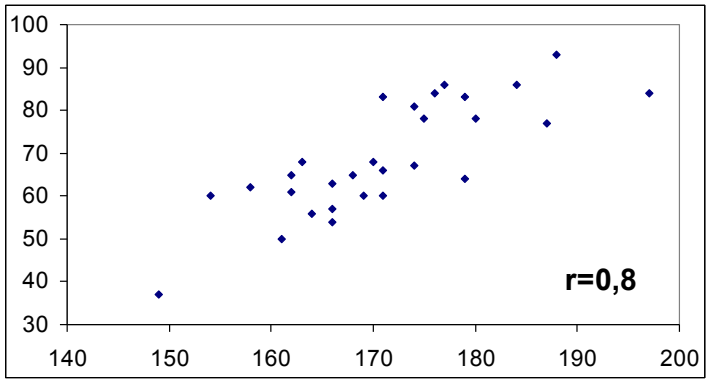


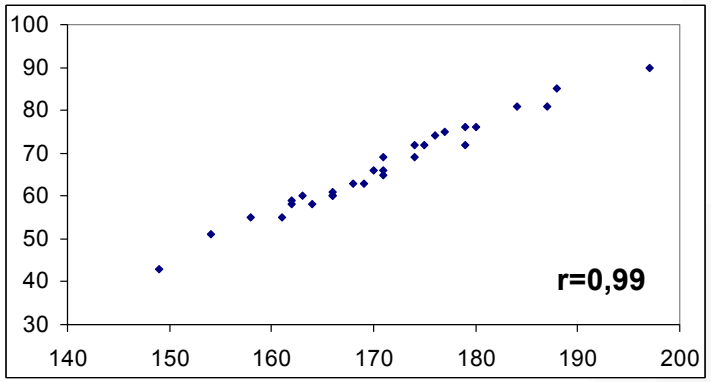
Ausencia de relación lineal.

Ejemplos. Correlaciones positivas

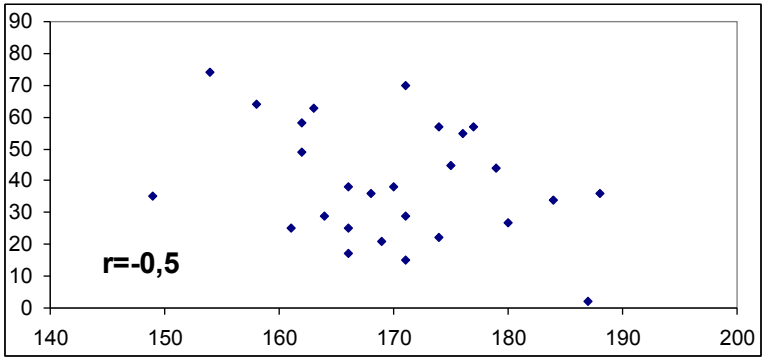


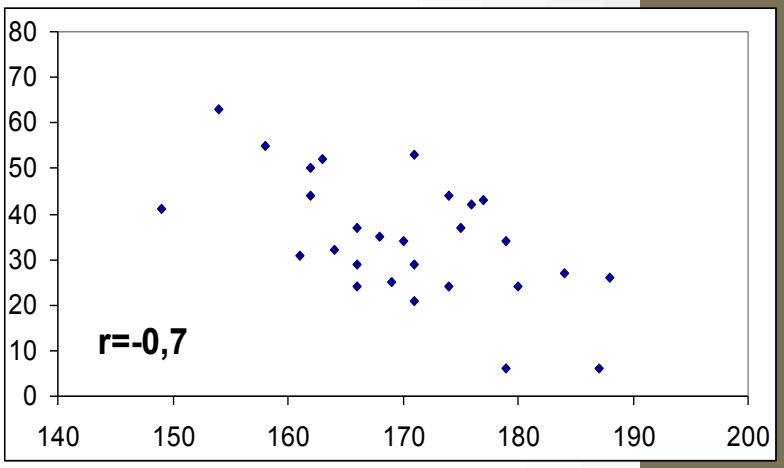


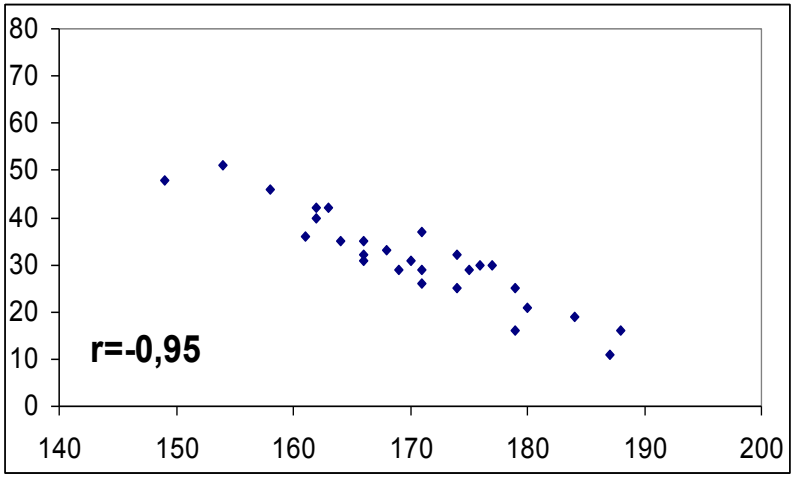


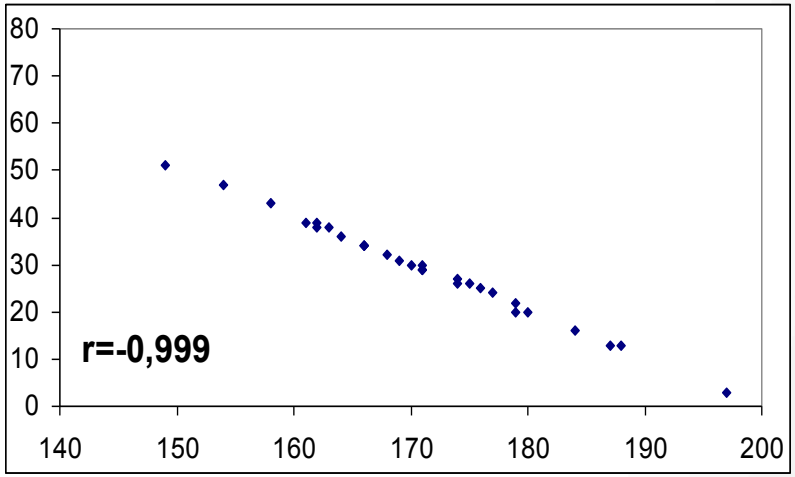


Ejemplos. Correlaciones negativas

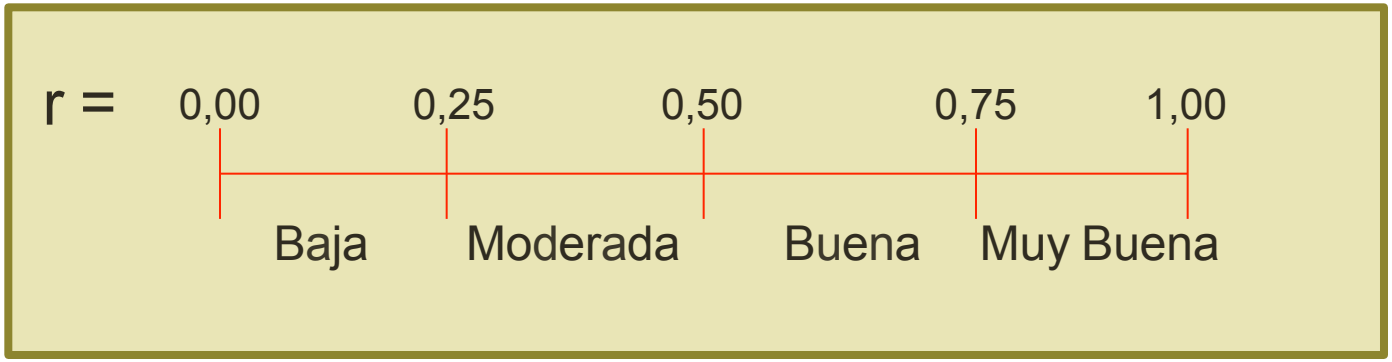








Magnitud de la correlación ‘r’



* Esta interpretación es similar cuando se tiene una correlación negativa.

Algunas propiedades de r

- Es adimensional
- Sólo toma valores entre -1 y 1
- Variables incorreladas ⇔r=0
- Relación lineal perfecta entre dos variables ⇔ r=+1 o r=-1
- Cuanto más cerca esté r de +1 o -1 mejor será el grado de relación lineal.
  - Siempre que no existan observaciones anómalas.

Otros coeficientes de correlación

- En el caso de variables ordinales, es posible analizar la correlación entre ellas:
  - Ro de Spearman
  - Tau de Kendall

Regresión Lineal Simple

- Sirve para predecir una medida en función de otra (o varias).

  Y= Variables dependiente (respuesta)

  X = Variable independiente (explicativa)

- La idea es encontrar una función que nos permita aproximar Y en función de X.

$$\hat{y} = \alpha + \beta x$$

α = ordenada en el origen (constante)

β = pendiente de la recta

Regresión Lineal Simple

$$\hat{y} = \alpha + \beta x$$

α = ordenada en el origen (constante)→ Es lo que vale y cuando x = 0.

β = pendiente de la recta → es lo que aumenta la y cuando x aumenta una unidad.

La variable explicativa (dependiente) influye sobre la variable respuesta (independiente).

Por cada aumenta de una unidad de la variable explicativa, la variable respuesta (independiente) aumenta(o disminuye) en b unidades.

Validación del modelo

$$\hat{y} = \alpha + \beta x$$

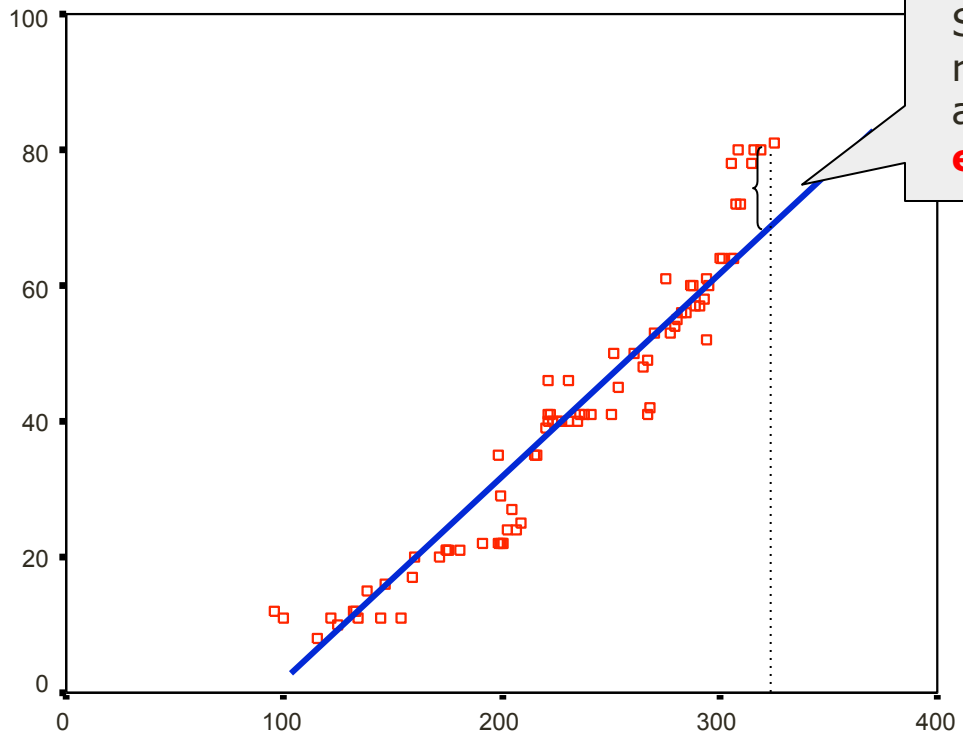
α = ordenada en el origen (constante)→ Es lo que vale y cuando x = 0.

β = pendiente de la recta → es lo que aumenta la y cuando x aumenta una unidad.

La variable explicativa (dependiente) influye sobre la variable respuesta (independiente).

Por cada aumenta de una unidad de la variable explicativa, la variable respuesta (independiente) aumenta(o disminuye) en b unidades.

Validación del modelo



VAR EXPLICATIVA

Siempre hay una diferencia entre el valor real de la variable respuesta y la estimación a partir de la ecuación de regresión:
el residuo

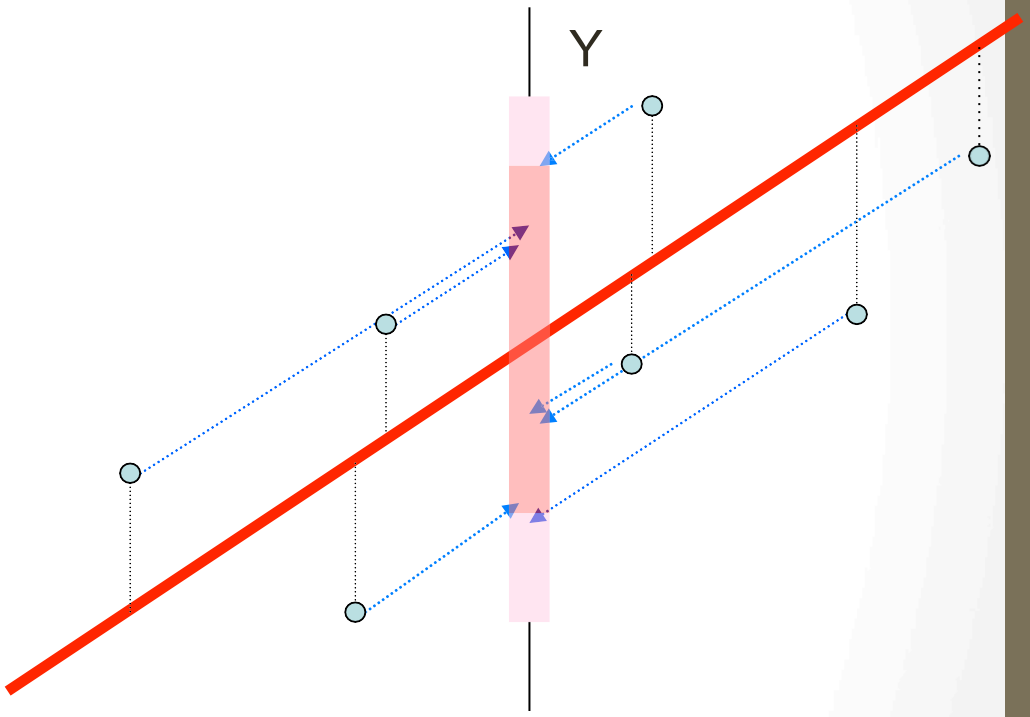
Validación

- Es preciso realizar una validación del modelo

La relación entre variables no es exacta.
?? Cuál es la mejor recta que sirve para predecir los valores de Y en función de X.
?? Que error cometemos con la aproximación

Recta de Regresión lineal simple Ajuste del modelo - Coeficiente R^2

$$R = 1 - \frac{S_e^2}{S_Y^2}$$



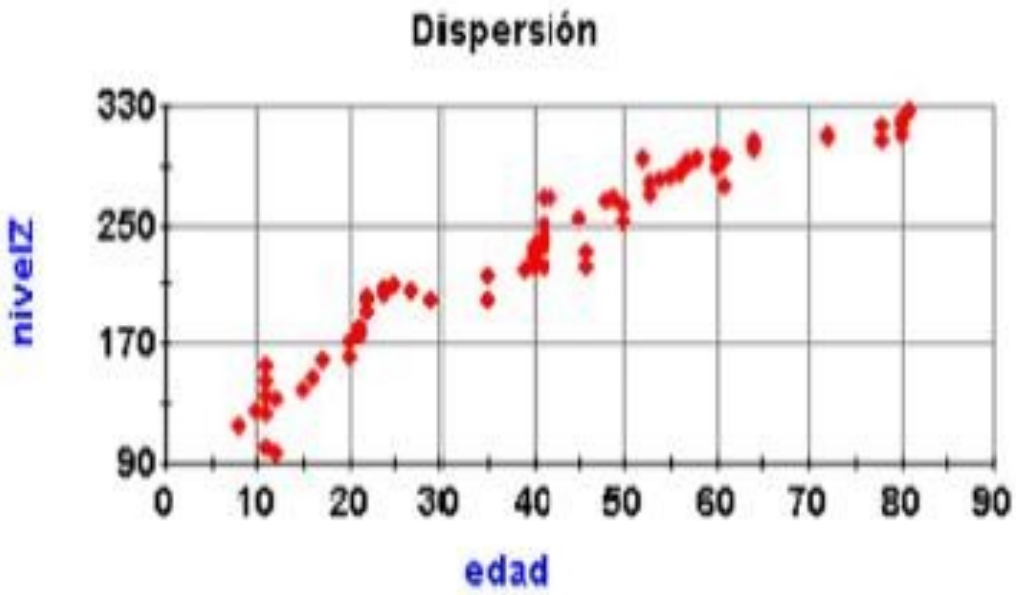
Propiedades del Coeficiente R^2

- ✓ La bondad de un ajuste de un modelo de regresión se mide usando el coeficiente de determinación R^2
- ✓ Cuando un ajuste es bueno, R^2 será cercano a 1 o 100%.
- ✓ Cuando un ajuste es malo R^2 será cercano a cero.
- ✓ A R^2 también se le denomina porcentaje de variabilidad explicado por el modelo de regresión.
- ✓ R^2 es una cantidad adimensional que sólo puede tomar valores en [0, 1]

Ejemplo:

De los 81 pacientes a los que se les ha administrado una medicación. A las 24 horas se determina la concentración sérica.

¿Depende la concentración sérica de la edad del paciente?



- El gráfico de dispersión muestra que los niveles son más elevados entre los pacientes de mayor edad
- La relación es de tipo lineal, por lo que una Regresión Lineal Simple será apropiada para modelizar la cuestión

Ejemplo: estimación de coeficientes

Análisis / x|y / Regresión Lineal Simple

nivel Z = 114.8 + 2.8 edad

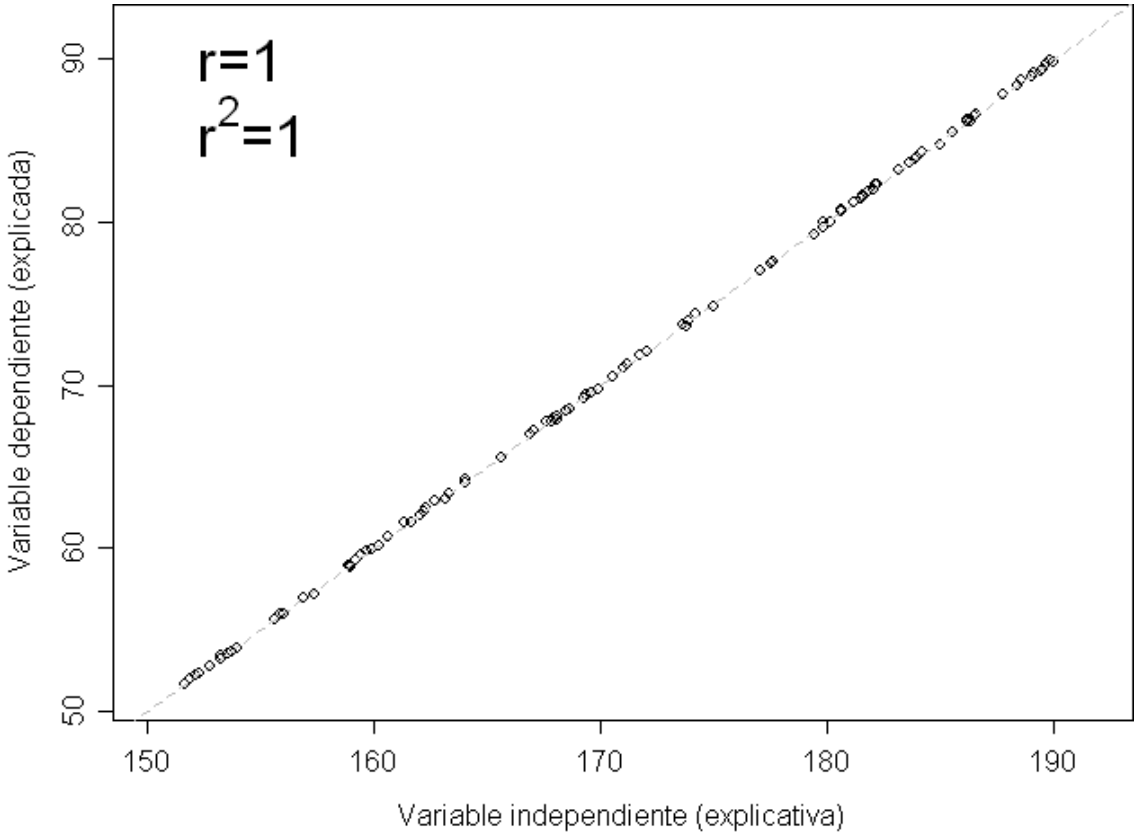
```
Número de Casos: 81

Modelo: Lineal
-----------------------------------------------------------------
Ecuación: nivelZ = 114.8159 + 2.8371 * edad
-----------------------------------------------------------------
              Coef.      E.E.      t-valor    p-valor
-----------------------------------------------------
 Ordenada     114.8159   4.2849    26.7955    0.0001
 Pendiente    2.8371     0.0941    30.1461    0.0001
-----------------------------------------------------


r de Pearson (coeficiente de correlación)  0.9592
r cuadrado (coeficiente de determinación)  92.00 %
Desviación Típica de los Residuos          17.2794
Rho de Spearman 0.9836   t-valor 48.4093   p-valor 0.0000
```

- Por cada año adicional de edad, la concentración en sangre aumenta 2,8 ml
- La edad explica el 92% de la varianza de los niveles. El ajuste es muy bueno.

Coeficientes de r y R^2 y diagrama de dispersión



Ejercicio 1.23

A partir de la base de datos LITIABIS.sav

- Calcule e interprete los coeficientes de correlación lineal de Pearson entre las variables EDAD, TALLA, PESO,y OBESIDAD.
- Construya un modelo de regresión lineal en función de los resultados del apartado anterior.

REPASO: descriptiva, gráficos, etc...

Opcional: Ej 1.6

En la Escuela de Enfermería se realizó una encuesta a 15 estudiantes en la que hacían constar el número de horas diarias dedicadas al estudio y la calificación obtenida en el último examen.

Los resultados ap ola adjunta:

Número Horas de estudio	Calificación
3.5	6
2.0	4.25
2.5	4.75
1.0	2.25
1.5	3.5
2.8	5.5
3.4	6
4.2	7.5
3.5	6.25
2.6	5.5
3.7	6.25
5.0	8.75
5.5	9.5
4.5	9.0
3.0	6

Opcional: Ej 1.6

i. Construya el diagrama de dispersión de los datos.

ii. Construya la tabla de distribución de frecuencias conjuntas para las variables número de horas de estudio, categorizada en menos de tres y más o igual a tres, y calificación, categorizada en menos de 5 y más o igual a 5.

iii. Calcule los porcentajes por filas y columnas de la tabla. Interprete los resultados. ¿Cuáles son los porcentajes de aprobados entre los alumnos que estudian más o igual a tres horas?. ¿Y entre los que lo hacen menos de 3 horas?. Construya gráficos para la distribución conjunta de ambas variables que le permitan visualizar los porcentajes por filas y por columnas