Estudio de Caso: Estudio Morfología[1]

Coeficiente de Correlación

Considere el archivo *"Estudio Morfología.sav".*

f) Determine si las variables estatura, peso y coeficiente intelectual están correlacionadas significativamente.

Para esto seleccione Analizar/Correlaciones/Bivariadas
Variables: ci, estatura y peso

Seleccionar [Correlación de Pearson], [Prueba de significación bilateral], [Marcar las correlaciones significativas].

Correlaciones

		Cociente intelectual	ESTATURA	PESO
Cociente intelectual	Correlación de Pearson	1	,081	,001
	Sig. (bilateral)	.	,325	,988
	N	149	149	148
ESTATURA	Correlación de Pearson	,081	1	,600**
	Sig. (bilateral)	,325	.	,000
	N	149	150	149
PESO	Correlación de Pearson	,001	,600**	1
	Sig. (bilateral)	,988	,000	.
	N	148	149	149

**. La correlación es significativa al nivel 0,01 (bilateral).

g) Interprete el coeficiente de correlación

5. Diagrama de Dispersión

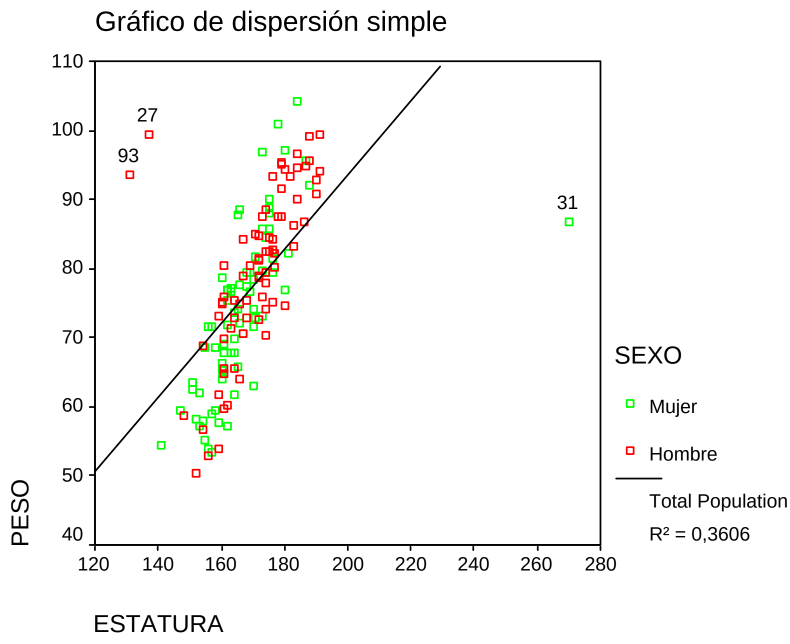
La forma de una relación se puede estudiar visualmente a partir de la nube de puntos generada en el Gráfico de Dispersión:

Seleccione Gráficos/Dispersión/Dispersión Simple

Eje Y:	peso
Eje X:	estatura
Establecer marcas por:	sexo
Etiquetar mediante:	iden

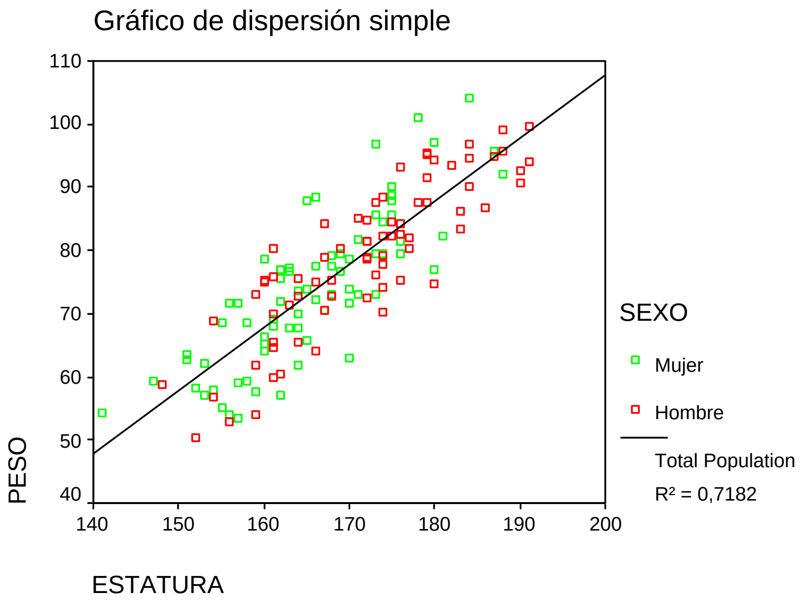
Edite la gráfica y ajuste la nube por una recta de regresión, muestre el R^2 e interprete.

[1] Caso elaborado por Sara Arancibia y Nelson Rodriguez



Descubra qué puntos están alejados de la nube y fíltrelos para volver a hacer el gráfico de dispersión, compare ahora el R^2 e interprete.

```
USE ALL.
COMPUTE filter_$=(iden ~= 27 & iden ~= 93 & iden ~= 31).
VARIABLE LABEL filter_$ 'iden ~= 27 & iden ~= 93 & iden ~= 31 (FILTER)'.
VALUE LABELS filter_$ 0 'No seleccionado' 1 'Seleccionado'.
FILTER BY filter_$.
EXECUTE .
```

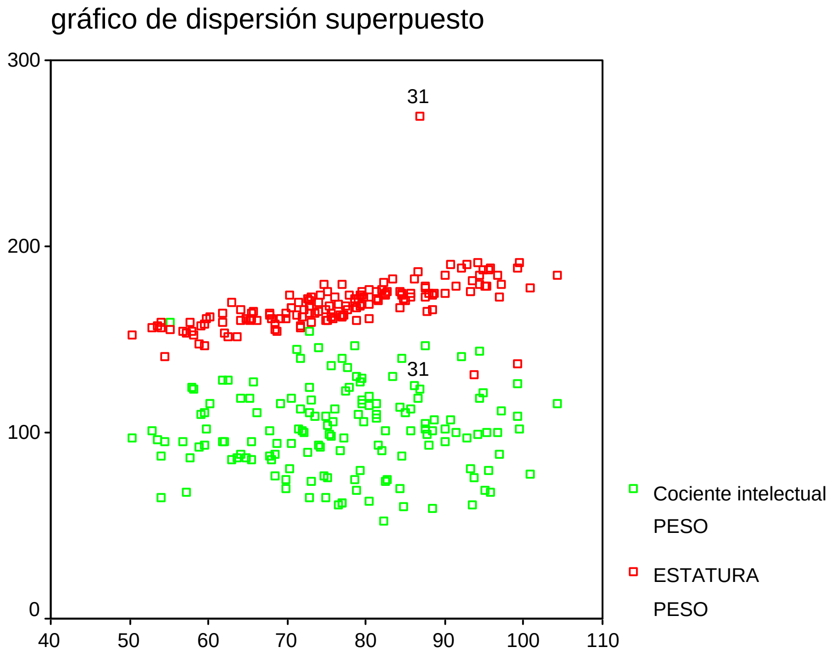


Se han filtrado los casos 27, 31 y 93

f) Realice un gráfico de dispersión **superpuesto**.

Con el tipo Superpuesto se pueden representar varias relaciones en una sola nubes de puntos.

Seleccione el par estatura - peso y el par ci - peso



g) Realice un gráfico de dispersión matricial.
El tipo matricial halla nubes de puntos separadas para todas las parejas de variables que se pueden obtener con las variables que se especifiquen.
Elegir el tipo matricial como tipo de gráfica de dispersión y seleccionar y transferir las variables a relacionar a variables en la matriz. Por ejemplo seleccionar y transferir ci, estatura y peso. Pulsar aceptar.

Como hay tres variables habrá tres parejas de relaciones (con cuatro variables habrá seis parejas).

Utilizando el archivo "*Estudio Morfología.sav*". responda las siguientes preguntas:

h) Determine el modelo que relaciona la estatura (X) y el peso (Y) e interprete R, R^2 y Error típico de estimación

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,847(a)	,718	,716	6,4593

a Variables predictoras: (Constante), ESTATURA

El coeficiente R mide la fuerza de asociación lineal entre estatura y peso, la cual es considerable.

El R^2 =0,718 indica que la variación en el peso se explica en un 71,8% por la variable estatura.

Una forma de estimar el error estándar del estimador es basándose en los residuos;

$$S_{Y,X} = \sqrt{\frac{\sum e^2}{n-2}}$$

El error estándar de la estimación es una medida de cuán inexacto podría ser la predicción y mide la dispersión con respecto a una recta promedio, denominada recta de regresión.

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	15309,683	1	15309,683	366,941	,000[a]
	Residual	6008,032	144	41,722		
	Total	21317,715	145			

a. Variables predictoras: (Constante), ESTATURA

b. Variable dependiente: PESO

Cuando se trata de una regresión simple la prueba ANOVA se reduce a la prueba individual Test T donde H_0: $\beta_1 = 0$

i) Estime la ecuación de regresión

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-92,138	8,816		-10,451	,000
	ESTATURA	,999	,052	,847	19,156	,000

a. Variable dependiente: PESO

$$\boldsymbol{Y = -92{,}13 + 0{,}999 \cdot X}$$

donde Y = Peso

X = Estatura

j) Pruebe la hipótesis nula H_0: β_1 = 0 para la estatura y el peso. ¿Existe una relación significativa entre el ingreso y el consumo?

En el SPSS ver la tabla "coeficientes" que resultó en el ejercicio b) y observar el valor del estadígrafo t (asociado a la pendiente de la regresión) y su nivel de significancia.

Dado que la sig < 0,01 se rechaza la hipótesis nula H_0: $\beta_1 = 0$. Concluimos entonces que existe una relación significativa entre peso y estatura.

k) Interprete la pendiente de la ecuación de regresión.

La pendiente de la recta b1 es el cambio que se produce en la variable dependiente (en promedio) por cada unidad de cambio en la variable independiente, es decir, por cada unidad de aumento en la estatura, en promedio el peso aumenta en 0,999 unidades.

Nota: Inferencias sobre la Pendiente.

A fin de poder utilizar una ecuación de regresión para efectos de estimación o predicción, primero debemos determinar si en la población parece existir una relación entre las dos variables o si la relación observada en la muestra pudo ocurrir por azar. En ausencia de toda relación en la población, por definición la pendiente de la línea de regresión de la población sería de cero β_1=0. En consecuencia, la hipótesis nula que se prueba usualmente es H_0: β_1=0. La hipótesis nula también puede formularse como una prueba de una cola, en cuyo caso la hipótesis alternativa no es simplemente que existe relación entre las dos variables, sino además que esta relación es de un tipo específico (directa o inversa).

Un valor hipotético de la pendiente se prueba calculando una estadística t y usando *n-2* grados de libertad. Es el proceso de inferencia se pierden dos grados de libertad porque en la ecuación de regresión se incluyen dos estimaciones paramétricas, b_0 y b_1. La fórmula estándar es:

$$t = \frac{b_1 - (\beta_1)o}{s_{b1}}$$

donde $s_{b1} = \dfrac{S_{Y,X}}{\sqrt{\sum X^2 - n\overline{X}^2}}$

Sin embargo, cuando, como ocurre por lo general, la hipótesis nula es que la pendiente es cero, la fórmula se simplifica y enuncia como

$$t = \frac{b_1}{s_{b1}}$$

El intervalo de confianza para la pendiente de la población β1, en el que los grados de libertad asociados con *t* son *n-2*, se elabora de la siguiente manera:

$$b_1 \pm t s_{b1}$$

Definición de grados de libertad: Los grados de libertad indican el número de valores "libres de variar" en la muestra que sirve de base al intervalo de confianza.

l) Determine el intervalo de confianza del 95% para β_1. Para esto seleccione Regresión lineal/ Estadísticos/Intervalos de confianza.

En la tabla de resultados "coeficientes" del SPSS observe los límites inferior y superior del intervalo de confianza para b_1 al 95%.

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	-92,138	8,816		-10,451	,000	-109,564	-74,712
	ESTATURA	,999	,052	,847	19,156	,000	,896	1,102

a. Variable dependiente: PESO

Se tiene que el intervalo de confianza de 95% para β_1 es 0,896 a 1,102

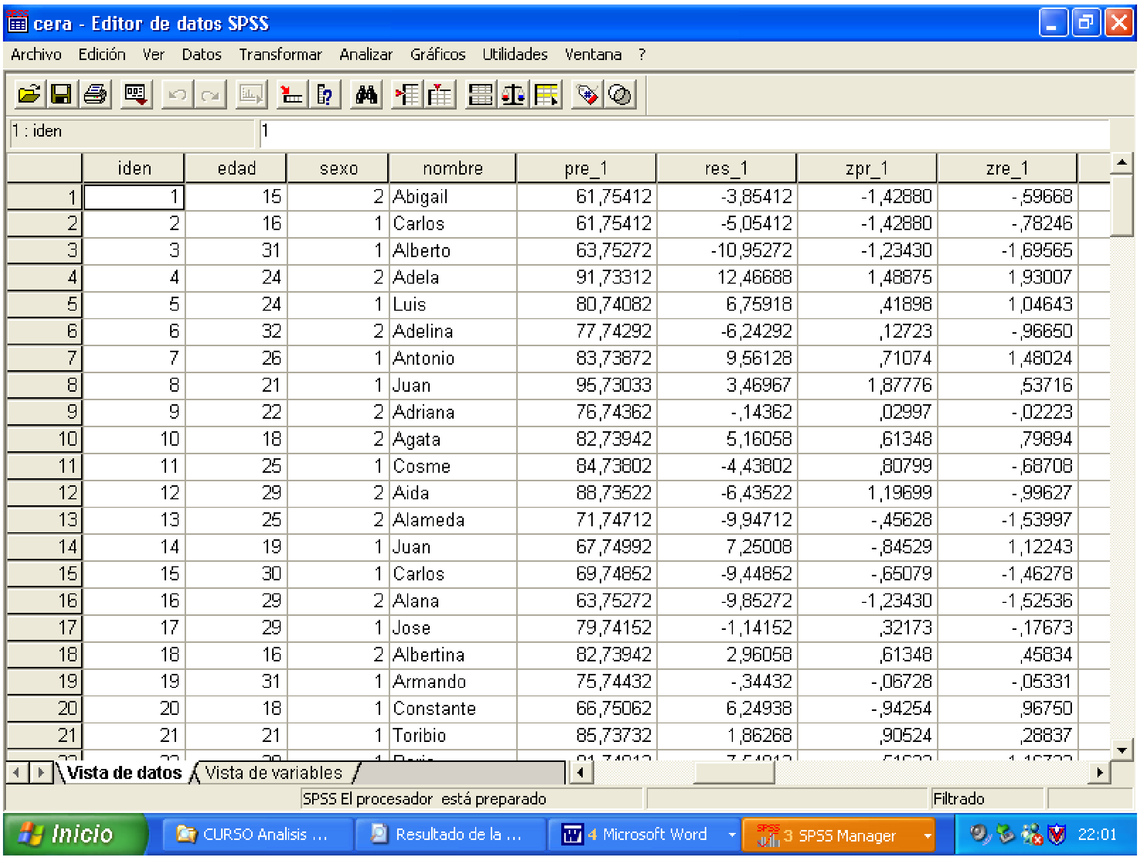
Así con cada unidad adicional de estatura, la cantidad de aumento promedio en el peso es de entre 0,896 y 1,102 con una confianza de 95%.

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	2,129	7,164		,297	,772	-13,834	18,092
	INGRESO	,861	,049	,984	17,596	,000	,752	,970

a. Variable dependiente: CONSUMO

m) Determine los valores pronosticados y los residuos usando la ecuación de regresión desarrollada. Compare los residuos obtenidos del SPSS. Para esto seleccione "Guardar" en el cuadro de diálogo "Regresión lineal" y en el cuadro de diálogo siguiente considere valores pronosticados no tipificados (es decir el valor que predice el modelo para la variable dependiente) y valores tipificados (transformación de cada valor pronosticado a su forma tipificada). Además considere residuos no tipificados (es decir, la diferencia entre un valor observado y el valor pronosticado del modelo) y los residuos tipificados.



	iden	edad	sexo	nombre	pre_1	res_1	zpr_1	zre_1
1	1	15	2	Abigail	61,75412	-3,85412	-1,42880	-,59668
2	2	16	1	Carlos	61,75412	-5,05412	-1,42880	-,78246
3	3	31	1	Alberto	63,75272	-10,95272	-1,23430	-1,69565
4	4	24	2	Adela	91,73312	12,46688	1,48875	1,93007
5	5	24	1	Luis	80,74082	6,75918	,41898	1,04643
6	6	32	2	Adelina	77,74292	-6,24292	,12723	-,96650
7	7	26	1	Antonio	83,73872	9,56128	,71074	1,48024
8	8	21	1	Juan	95,73033	3,46967	1,87776	,53716
9	9	22	2	Adriana	76,74362	-,14362	,02997	-,02223
10	10	18	2	Agata	82,73942	5,16058	,61348	,79894
11	11	25	1	Cosme	84,73802	-4,43802	,80799	-,68708
12	12	29	2	Aida	88,73522	-6,43522	1,19699	-,99627
13	13	25	2	Alameda	71,74712	-9,94712	-,45628	-1,53997
14	14	19	1	Juan	67,74992	7,25008	-,84529	1,12243
15	15	30	1	Carlos	69,74852	-9,44852	-,65079	-1,46278
16	16	29	2	Alana	63,75272	-9,85272	-1,23430	-1,52536
17	17	29	1	Jose	79,74152	-1,14152	,32173	-,17673
18	18	16	2	Albertina	82,73942	2,96058	,61348	,45834
19	19	31	1	Armando	75,74432	-,34432	-,06728	-,05331
20	20	18	1	Constante	66,75062	6,24938	-,94254	,96750
21	21	21	1	Toribio	85,73732	1,86268	,90524	,28837

Observación:
Si en la regresión lineal queremos llevar a cabo inferencias y partimos de los estadísticos obtenidos en la muestra, deberemos tener en cuenta una serie de requisitos:

- Normalidad e igualdad de las varianzas en la variable dependiente (Y) del modelo para valores fijos de la independiente o independientes del mismo X.
- Independencia de las observaciones
- Linealidad en la relación entre las variables.

n) Considere "Gráficos" del cuadro de diálogo "Regresión lineal" para realizar los siguientes gráficos:

(i) Los residuos tipificados ZRESID frente a los valores pronosticados tipificados ZPRED para contrastar la igualdad de las varianzas.

Nota: Si no hay ningún patrón sistemático claramente definido en los datos y los residuales fluctúan aleatoriamente alrededor de la recta que corresponde a la media de los mismos y de valor cero, podemos concluir que se cumple el requisito de linealidad en la relación entre las variables. Este gráfico puede igualmente servirnos para contrastar hasta qué punto el principio de igualdad de varianzas puede o no ser violado por los datos. Si la variabilidad de los residuales a lo largo de los valores predichos es más o menos constante, podemos concluir que se cumple la igualdad de varianzas. No en caso contrario.

(ii) Los residuos tipificados-gráfico de prob. normal.

<u>Nota</u>: El gráfico de residuos tipificados de prob. normal se usa para comprobar la normalidad. Si la variable se distribuye normalmente los puntos representados forman una línea recta diagonal

Gráfico P-P normal de regresión Residuo tipificado

Variable dependiente: PESO

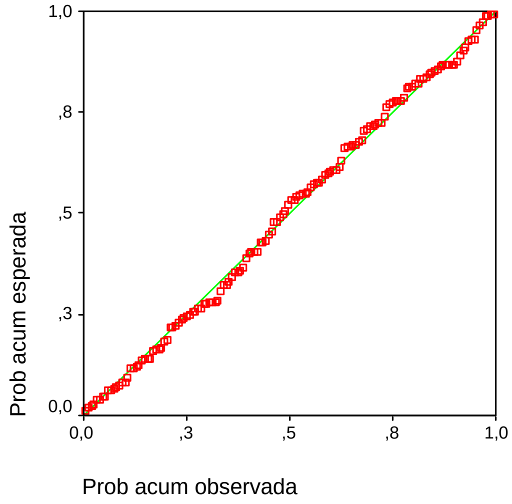
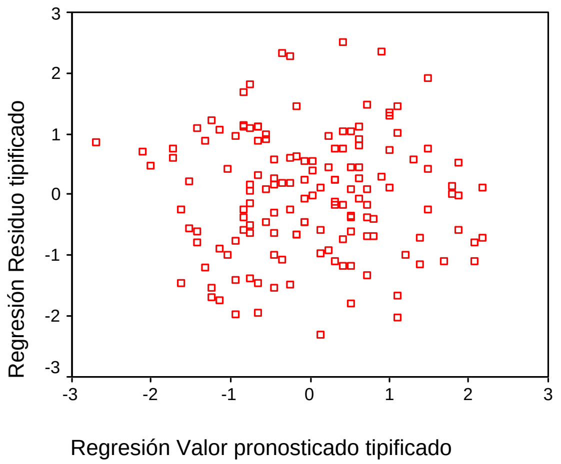


Gráfico de dispersión

Variable dependiente: PESO



Pruebas de normalidad

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Standardized Residual	,044	146	,200*	,992	146	,601

*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors

después del año actual.