

# ESTADÍSTICA

## PRÁCTICA 6. El problema de la dependencia entre variables medibles.

**Nombre:** Pedro Giménez Aldeguer

**Grupo Prácticas:** 02

**DNI:** 15419933C

**Email:** pedrogial@gmail.com

**Fecha:** 18/12/2017

# Índice

• Ejercicio 1:	3
• a)	3
• b)	4
• c)	5
• d)	7
• e)	9
• f)	10
• Ejercicio 2:	10
• a)	10
• b)	11
• c)	12
• d)	13

1. Se realiza un estudio para investigar la relación entre el nivel de humedad del suelo y la tasa de mortalidad en lombrices de tierra. La tasa de mortalidad, Y, es la proporción de lombrices de tierra que mueren tras un periodo de dos semanas. El nivel de humedad, X, viene medido en milímetros de agua por centímetro cuadrado de suelo. Se obtuvieron los siguientes datos:

X	0.00	0.00	0.00	0.31	0.31	0.31	0.63	0.63	0.63	0.94	0.94	0.94	1.26	1.26	1.26
Y	0.5	0.4	0.5	0.2	0.3	0.3	0.0	0.1	0.0	0.1	0.2	0.1	0.6	0.5	0.4

En SPSS:

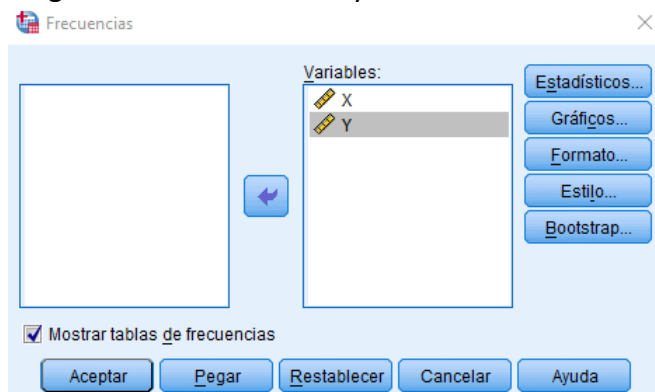
	X	Y
1	,00	,5
2	,00	,4
3	,00	,5
4	,31	,2
5	,31	,3
6	,31	,3
7	,63	,0
8	,63	,1
9	,63	,0
10	,94	,1
11	,94	,2
12	,94	,1
13	1,26	,6
14	1,26	,5
15	1,26	,4

Calcular:

**a) La media y la varianza.**

Para saber la media y la varianza vamos a la opción “Analizar”, “Estadístico descriptivo”, “Frecuencias...”.

Elegimos las dos variables y le damos a “Estadísticos...”:



Aquí seleccionamos la Media y la Varianza y le damos a continuar:

Esta es nuestra tabla con los datos:

Estadísticos			
		X	Y
N	Válido	15	15
	Perdidos	0	0
Media		,6280	,280
Varianza		,213	,039

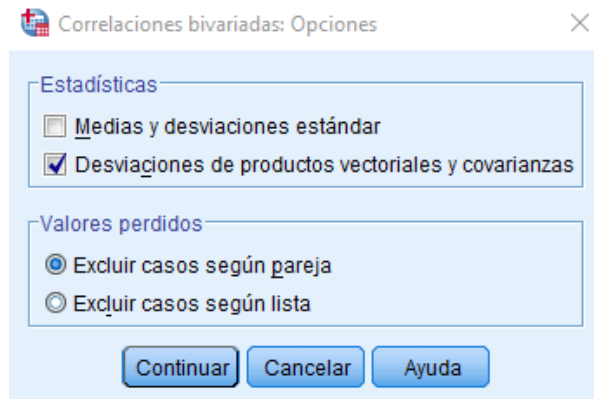
La Media de X es 0,6280 y la de Y es 0,270.

La Varianza de X es 0,213 y la de Y es 0,039.

#### b) La covarianza y el coeficiente de correlación.

Para esto te vas a la opción de “Analizar”, “Correlaciones” y “Bivariadas...”. Seleccionamos las variables y la opción de Pearson, además de darle a “Opciones...”:

En las opciones seleccionamos Desviaciones de productos vectoriales y covarianzas para poder sacar la covarianza de X e Y:



Este es el resultado:

### → Correlaciones

Correlaciones		X	Y
X	Correlación de Pearson	1	-,048
	Sig. (bilateral)		,866
	Suma de cuadrados y productos vectoriales	2,977	-,061
	Covarianza	,213	-,004
	N	15	15
Y	Correlación de Pearson	-,048	1
	Sig. (bilateral)	,866	
	Suma de cuadrados y productos vectoriales	-,061	,544
	Covarianza	-,004	,039
	N	15	15

La Correlación es -0,048 y la Covarianza -0,004.

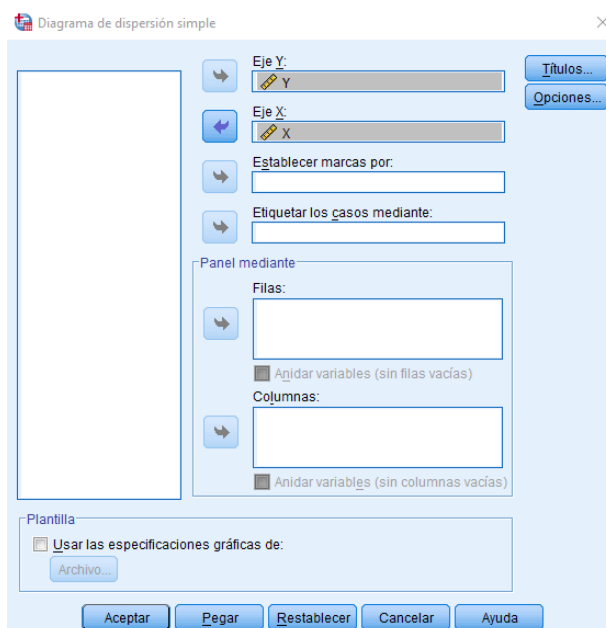
### c) Construir el diagrama de dispersión de las variables X e Y. ¿Sugieren los datos algún tipo de relación?

Para esto debemos ir a “Gráficos”, “Cuadro de diálogos antiguos”, “Dispersión...”.

Después, marcar la opción “Simple” y pulsamos “Definir”:

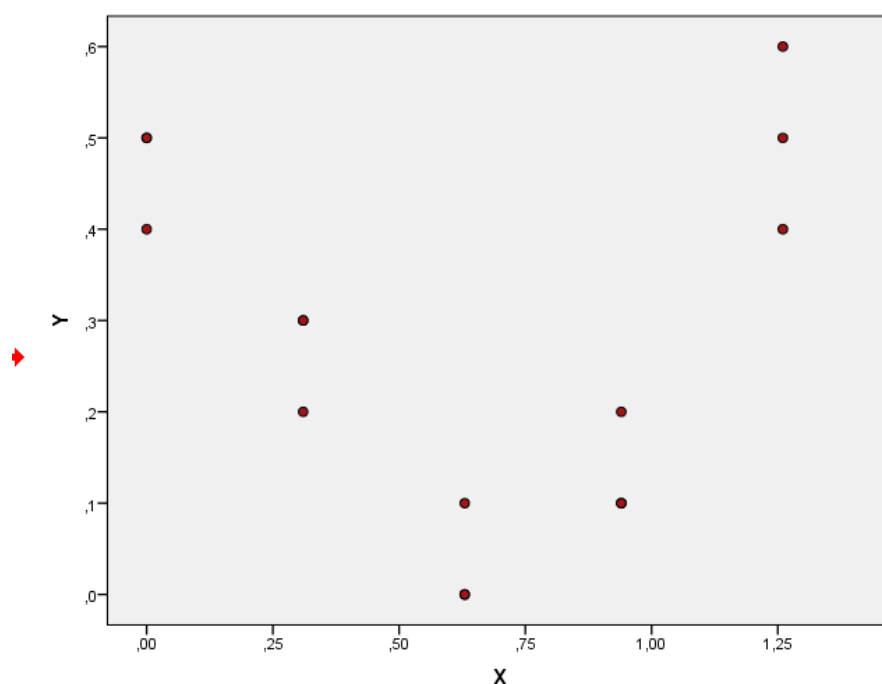


En el cuadro de diálogo obtenido seleccionamos las variables para los ejes, el tipo de gráfico y pulsamos “Aceptar”:



Este es el gráfico resultante:

**Gráfico**



Los datos están muy dispersos entre ellos, la única relación que tienen es que para cada valor de X contiene al menos dos puntos, pero en distinto valor de Y.

**d) Calcular las rectas de regresión.**

La secuencia de comandos “Analizar”, “Regresión”, “Estimación curvilínea...” abre el cuadro de diálogo, en él trasladamos la variable dependiente e independiente a las casillas correspondientes y marcamos Incluir constante en la ecuación y Lineal (si queremos la representación gráfica marcamos también Representar los modelos).

Recta de regresión de Y sobre X:

(Pasos de la explicación anterior, para el caso de Y sobre X)

Esta es la tabla resultante:

**Resumen de modelo y estimaciones de parámetro**

Variable dependiente: Y

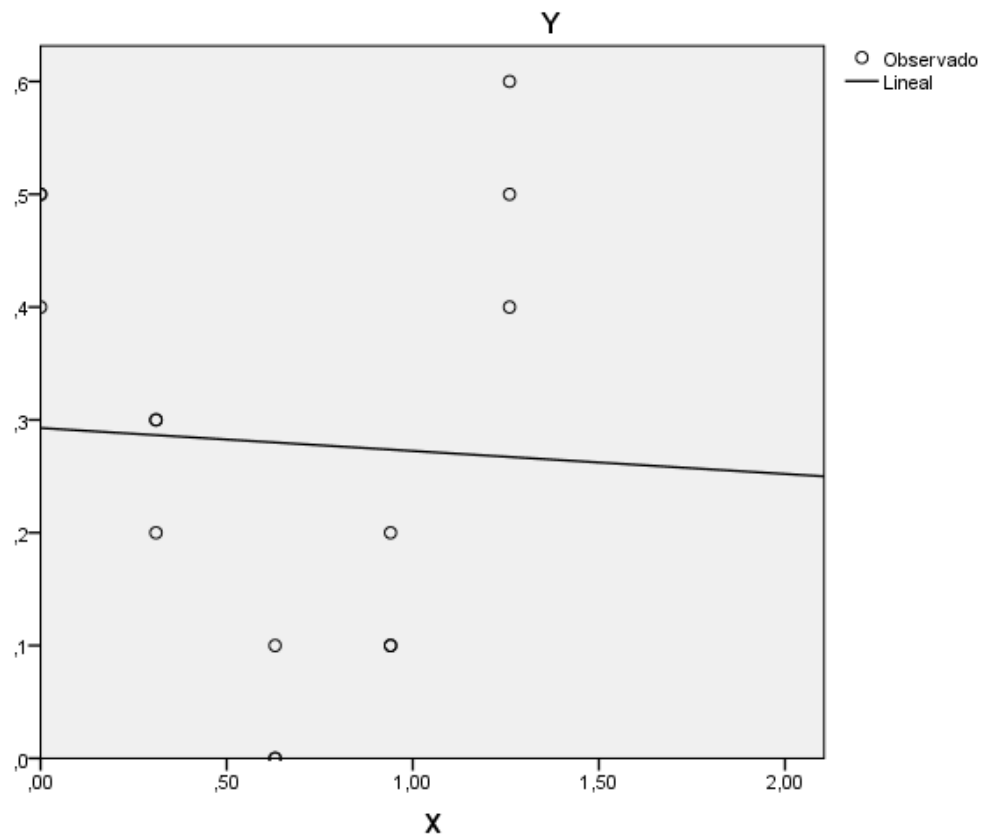
Ecuación	Resumen del modelo					Estimaciones de parámetro	
	R cuadrado	F	df1	df2	Sig.	Constante	b1
Lineal	,002	,030	1	13	,866	,293	-,020

La variable independiente es X.

Por lo que la ecuación de la recta de regresión de Y sobre X es:  
(Se coge b1 y la constante)

$$\underline{\underline{Y = -0,2x + 0,293}}$$

Y esta es la recta de regresión en el gráfico:



Recta de regresión de X sobre Y:  
(Pasos de la explicación anterior, para el caso de X sobre Y)

Estimación curvilínea

Dependientes: X

Independiente: Variable: Y

Etiquetas de caso: ☐ Incluir la constante en la ecuación ☒ Representar los modelos

Modelos: ☒ Lineal ☐ Cuadrático ☐ Compuesto ☐ Crecimiento ☐ Logarítmico ☐ Cúbico ☐ S ☐ Exponencial ☐ Inverso ☐ Potencia ☐ Logística

Límite superior:

☐ Ver tabla de ANOVA

Aceptar Pegar Restablecer Cancelar Ayuda

Guardar...



Esta es la tabla resultante:

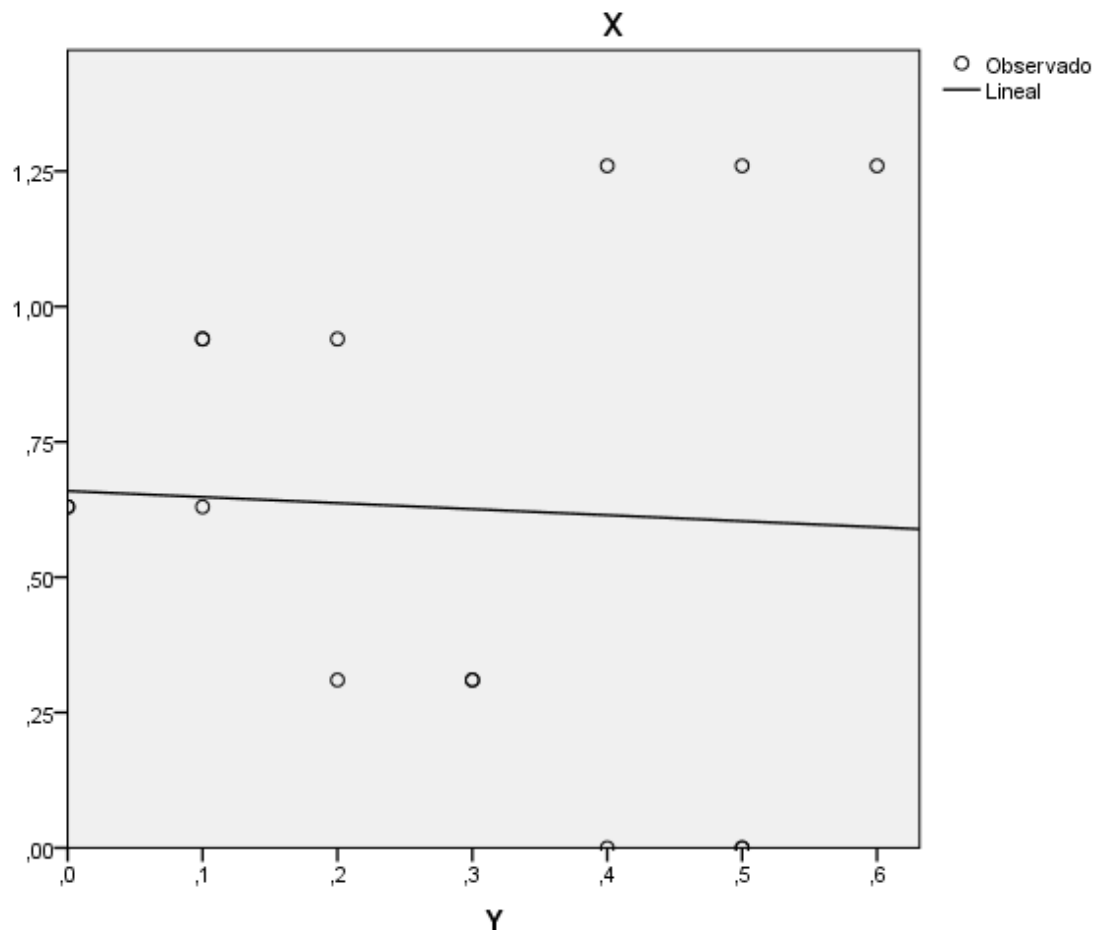
Resumen de modelo y estimaciones de parámetro							
Variable dependiente: X							
Ecuación	Resumen del modelo					Estimaciones de parámetro	
	R cuadrado	F	df1	df2	Sig.	Constante	b1
Lineal	,002	,030	1	13	,866	,659	-,111

La variable independiente es Y.

Por lo que la ecuación de la recta de regresión de X sobre Y es:  
(Se coge b1 y la constante)

$$X = -0,111y + 0,659$$

Y esta es la recta de regresión en el gráfico:



#### e) Interpretación.

Estas rectas presentan un mal ajuste ya que las variables están muy dispersas a la recta.

Además, la Correlación de Pearson que es -0,2 entonces no presenta una fuerte dependencia lineal decreciente, es independiente. Si las variables fueran dependientes decrecientes, la correlación debería estar cerca del valor -1. El ajuste lineal no es apropiado.

f) En una nueva medición el nivel de humedad X ha medido 1.57 ¿qué tasa de mortalidad Y se puede esperar?

Coges la ecuación de la recta de regresión de Y sobre X calculada antes, y calculas la tasa de mortalidad Y:

$$(Y = -0,2x + 0,293)$$

$$Y = -0,2 \cdot 1,57 + 0,293 // Y = 0,261$$

Se puede esperar una tasa de mortalidad de 0,261.

2. La tabla siguiente representa los datos correspondientes a 10 pruebas donde se ha estudiado el volumen de desgaste de una pieza (Y) dependiendo de la viscosidad del aceite (X):

X (aceite)	1.6	9.4	15.5	20.0	22.0	35.5	43.0	40.5	33.0	30.5
Y (desg.)	240	181	193	155	172	110	113	75	94	88

En SPSS:

	X_aceite	Y_desgaste
1	1,6	240
2	9,4	181
3	15,5	193
4	20,0	155
5	22,0	172
6	35,5	110
7	43,0	113
8	40,5	75
9	33,0	94
10	30,5	88

Se pide:

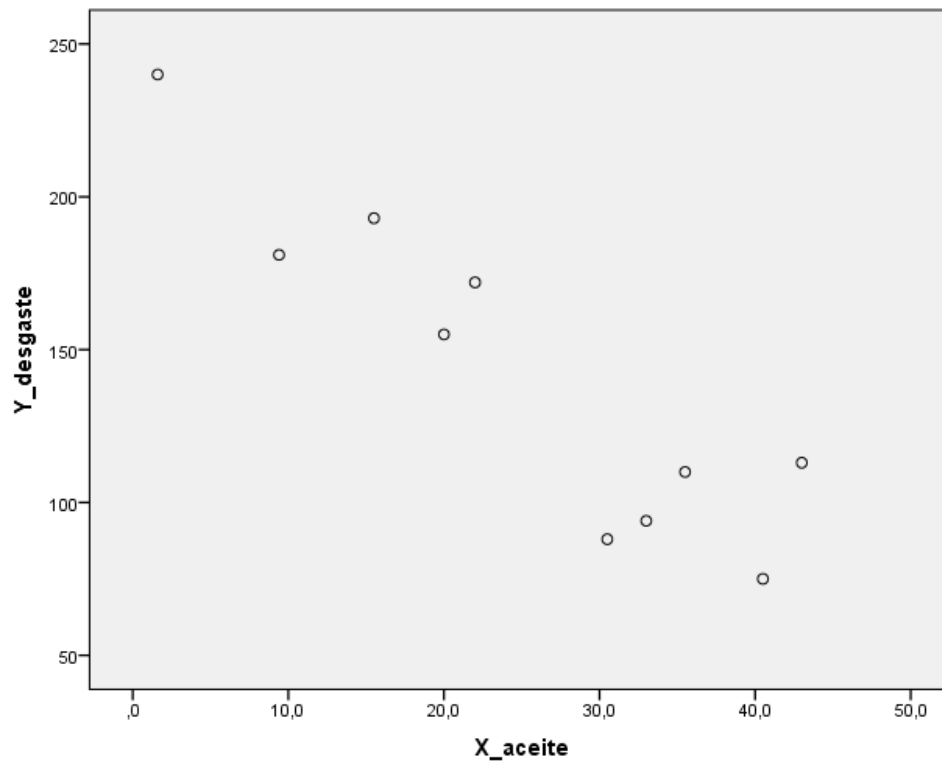
a) Construye el gráfico de dispersión (nube de puntos) de los datos. ¿Parece plausible ajustar una recta de regresión? ¿Cómo debe salir el coeficiente de correlación? Razona la respuesta.

(La explicación de la creación del gráfico de dispersión se encuentra en el ejercicio 1)

Si parece plausible ajustar una recta de regresión, ya que los puntos no están muy dispersos.

La correlación debe salir cercano a -1 ya que los puntos no están bastante dispersos y va decreciendo.

Este es el gráfico de dispersión resultante:



**b) Calcular la covarianza existente entre ambas variables, así como el coeficiente de correlación.**

(Explicación de la creación en el ejercicio 1)

Tabla resultante:

Correlaciones		X_aceite	Y_desgaste
X_aceite	Correlación de Pearson	1	-,918**
	Sig. (bilateral)		,000
	Suma de cuadrados y productos vectoriales	1683,820	-6118,700
	Covarianza	187,091	-679,856
	N	10	10
Y_desgaste	Correlación de Pearson	-,918**	1
	Sig. (bilateral)	,000	
	Suma de cuadrados y productos vectoriales	-6118,700	26368,900
	Covarianza	-679,856	2929,878
	N	10	10

\*\* . La correlación es significativa en el nivel 0,01 (2 colas).

La Correlación de Pearson es -0,918 y la covarianza -679,856.

**c) Calcular las dos rectas de regresión.**

(Explicación de la creación de las rectas de regresión en el ejercicio 1)

Recta de regresión de Y sobre X:

**Resumen de modelo y estimaciones de parámetro**

Variable dependiente: Y\_desgaste

Ecuación	Resumen del modelo					Estimaciones de parámetro	
	R cuadrado	F	df1	df2	Sig.	Constante	b1
Lineal	,843	43,020	1	8	,000	233,309	-3,634

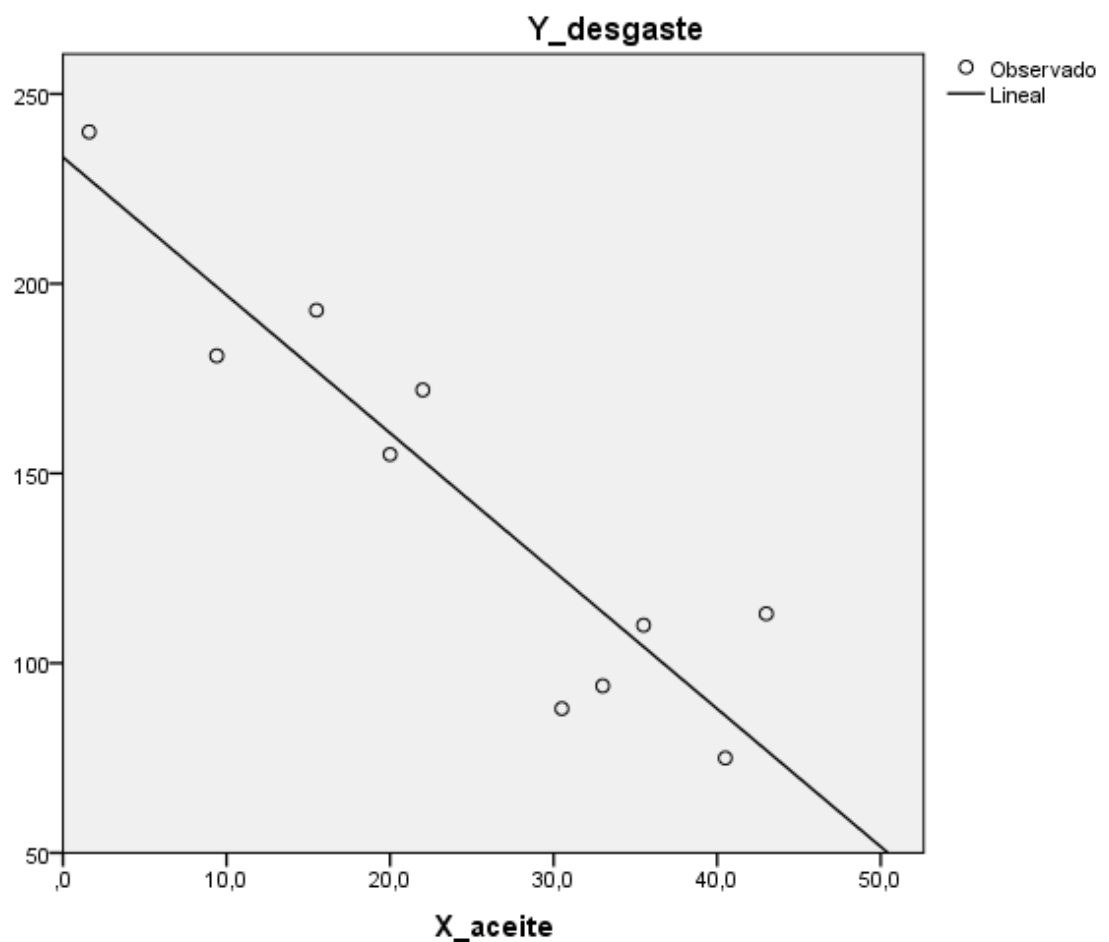
La variable independiente es X\_aceite.

Por lo que la ecuación de la recta de regresión de Y sobre X es:

(Se coge b1 y la constante)

$$Y = -3,634x + 233,309$$

Y esta es la recta de regresión en el gráfico:



Recta de regresión de X sobre Y:

**Resumen de modelo y estimaciones de parámetro**

Variable dependiente: X\_aceite

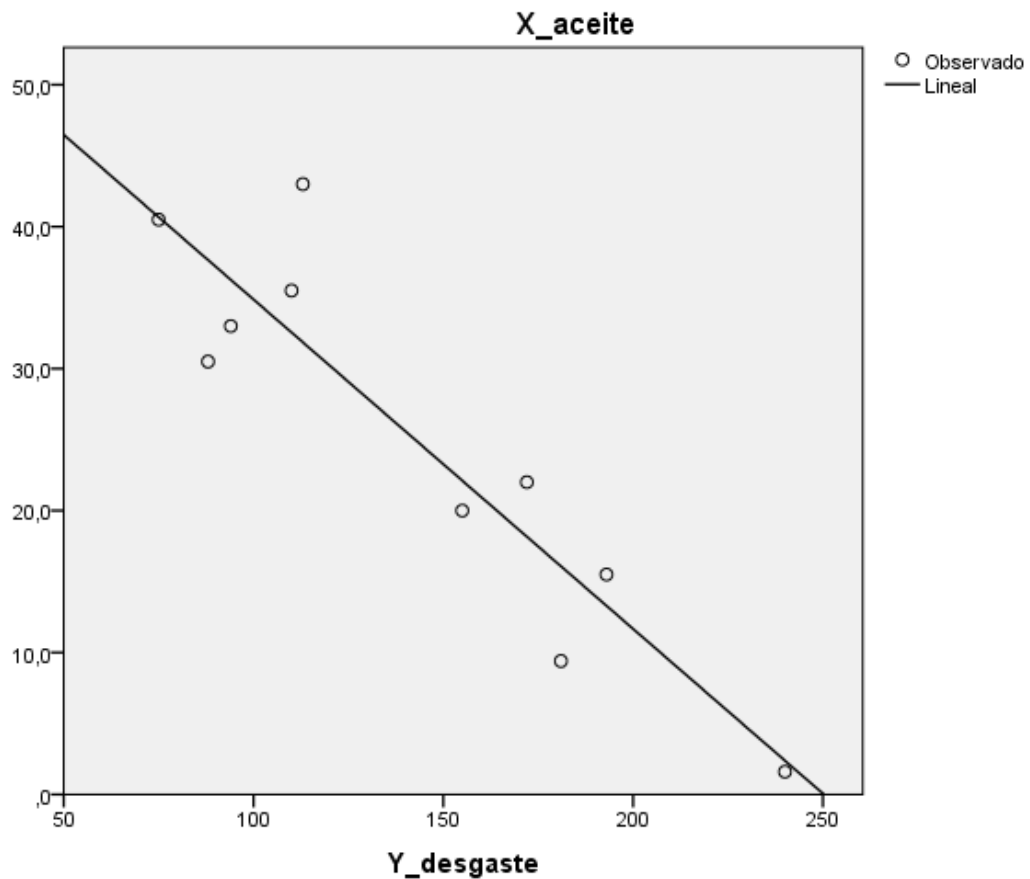
Ecuación	Resumen del modelo					Estimaciones de parámetro	
	R cuadrado	F	df1	df2	Sig.	Constante	b1
Lineal	,843	43,020	1	8	,000	58,073	-,232

La variable independiente es Y\_desgaste.

Por lo que la ecuación de la recta de regresión de X sobre Y es:  
(Se coge b1 y la constante)

$$\underline{X = 58,073y + -0232}$$

Y esta es la recta de regresión en el gráfico:



**d) ¿Cuál sería la predicción sobre el desgaste para una viscosidad de 45?**

Para saber la predicción sobre el desgaste se va a usar la ecuación de la recta de regresión de Y sobre X.

$$(Y = -3,634x + 233,309)$$

$$Y = -3,634 \cdot 45 + 233,309 \text{ // } Y = 69,779$$

La predicción sobre el desgaste para viscosidad de 45 es de 69,779.