



Análisis de los resultados de la evaluación del aprendizaje de procedimientos de análisis descriptivo en ingeniería informática

Apellidos, nombre	Capilla Romá, Carmen ¹ (ccapilla@eio.upv.es)
Departamento	¹ Estadística e Investigación Operativa Aplicadas y Calidad
Centro	Universitat Politècnica de València

1 Resumen de las ideas clave

En este artículo vamos a presentar los procedimientos estadísticos adecuados para analizar los resultados académicos de las evaluaciones de Estadística Descriptiva. Puede reutilizarse en contextos educativos distintos. El examen primer parcial también tiene una pregunta relacionada con estos conceptos. Se procede en primer lugar (Tabla 1) al análisis descriptivo univariante de cada una de las evaluaciones, mediante el cálculo de los parámetros descriptivos y representaciones gráficas con histograma, polígono de frecuencias, y diagrama Box&Whisker. A continuación se estudia la relación entre los resultados de evaluación con estadística descriptiva bidimensional. Finalmente se aplican modelos de regresión lineal simple para estudiar dichas relaciones. Se utiliza el programa Statgraphics para obtener todos los análisis.

Procedimientos que se aplican
1. Estadística descriptiva unidimensional
2. Estadística descriptiva bidimensional
3. Regresión lineal simple

Tabla 1. Análisis estadísticos de las evaluaciones.

2 Introducción

La aplicación de la Estadística permite por un lado obtener datos que tengan información representativa sobre un problema. Por otro lado el análisis estadístico posibilita extraer de dichos datos la información relevante para la toma de decisiones que implica el problema.

La primera etapa en el análisis de un conjunto de datos o muestra, consiste habitualmente en la aplicación de métodos descriptivos sencillos. El tratamiento descriptivo de las observaciones extrae de las mismas las características y regularidades más relevantes. Sintetiza la información con parámetros muestrales y representaciones gráficas. En este primer paso de análisis exploratorio, no se extrapolan conclusiones de los datos a la población de donde proceden. Esta extrapolación es parte de la inferencia estadística.

Los datos que se analizan son los resultados de tres evaluaciones. La muestra está formada por 40 alumnos que cursaban estadística en ingeniería informática. Son evaluaciones de la parte de estadística descriptiva unidimensional. La variable aleatoria es tridimensional con tres componentes numéricas continuas, que oscilan en el intervalo $[0, 10]$. Una evaluación se realizó trabajando en grupo con un compañero en el aula informática. Consistió en una prueba objetiva sobre la interpretación de análisis descriptivos univariantes y tablas de frecuencias bivariantes. Otra evaluación se realizó durante la sesión de seminario. En ella trabajaron en grupo con otro compañero. Calcularon parámetros descriptivos de posición y dispersión, y construyeron un diagrama Box&Whisker. La tercera

evaluación fue la pregunta del primer parcial sobre estadística descriptiva unidimensional, realizada individualmente.

En la primera etapa del análisis de estos datos, se estudia por separado cada componente mediante los métodos adecuados de estadística descriptiva unidimensional. Estos son: cálculo de parámetros de posición, dispersión y asimetría y curtosis. Se complementa el tratamiento con las representaciones gráficas de histograma y diagrama Box&Whisker, para estudiar la forma de la distribución de frecuencias y la posible existencia de datos anómalos.

En la segunda fase se investigan las relaciones bidimensionales entre las evaluaciones, utilizando el diagrama de dispersión y la correlación. Finalmente en la tercera etapa se aplican métodos de inferencia. Concretamente se realizan los ajustes de regresión lineal simple para analizar con mayor detalle las correlaciones. Se aplican contrastes de hipótesis sobre la ordenada y pendiente de las rectas estimadas.

3 Objetivos

Una vez que el alumno estudie con detenimiento este documento, será capaz de:

- Realizar e interpretar adecuadamente análisis descriptivos univariantes de variables numéricas continuas.
- Aplicar métodos estadísticos bivariantes para explorar relaciones entre variables continuas.
- Realizar y validar modelos de regresión lineal simple para estimar relaciones bivariantes.
- Aplicar e interpretar los resultados del software estadístico Statgraphics para realizar los análisis planteados.

Los estudiantes dispondrán de un conjunto de herramientas que les permitirán autoevaluarse, ya que realizarán durante el curso las mismas tres evaluaciones que analizan en la muestra de 40 alumnos. Conocerán la manera de analizar sus propios resultados de evaluación.

4 Desarrollo

Los **conocimientos previos** necesarios para abordar el estudio son:

- Conceptos básicos de **población y muestra**.
- **Variable aleatoria**: Una característica es aleatoria cuando puede constatarse en cada individuo de la población y presenta **variabilidad**. Si se expresa numéricamente en una escala continua es una variable de tipo **continuo**. Por ejemplo el resultado de una evaluación es una variable aleatoria continua en el intervalo $[0,10]$.
- **Histograma**: Representación gráfica en la que en el eje de abscisas se constatan los valores de la variable agrupados en tramos. Sobre cada tramo se levanta una barra de altura proporcional a la frecuencia (absoluta o relativa) observada en cada tramo.
- **Parámetros de posición**: Media, mediana y cuartiles.

- **Parámetros de dispersión:** Recorrido, varianza, desviación típica e intervalo intercuartílico.
- **Parámetros de asimetría y curtosis:** El de asimetría cuantifica si los datos son simétricos respecto a su media. El segundo determina si las frecuencias de valores alejados de la media son las que corresponden a la distribución Normal.
- **Diagrama Box&Whisker:** Representa la media, mediana, cuartiles y los valores mínimo y máximo de los datos.
- **Diagrama de dispersión:** Representación gráfica en la que en el eje de abscisas va el valor de la primera componente de una variable bidimensional, y en el de ordenadas la segunda componente.
- **Correlación r:** es la covarianza dividida por el producto de las desviaciones típicas. Mide el grado de relación lineal pero de forma adimensional.
- **Recta de regresión o modelo de regresión lineal simple:** Ajuste de mínimos cuadrados que calcula la ordenada y la pendiente de una recta, que permite calcular la media de una componente de la variable bidimensional en función del valor de la otra componente. Se aplican los contrastes t de Student sobre la significación estadística de estos dos parámetros.

4.1 Análisis descriptivo univariante

Para obtener este análisis se calculan los parámetros de posición, dispersión y asimetría y curtosis. Se representan el histograma y el diagrama Box&Whisker. Con un programa como Statgraphics, una vez abierto el fichero (Figura 1), la opción es **Describe....Numerical Data....One Variable Analysis.**

pract_desc: nota de la práctica de estadística descriptiva en el aula informática.

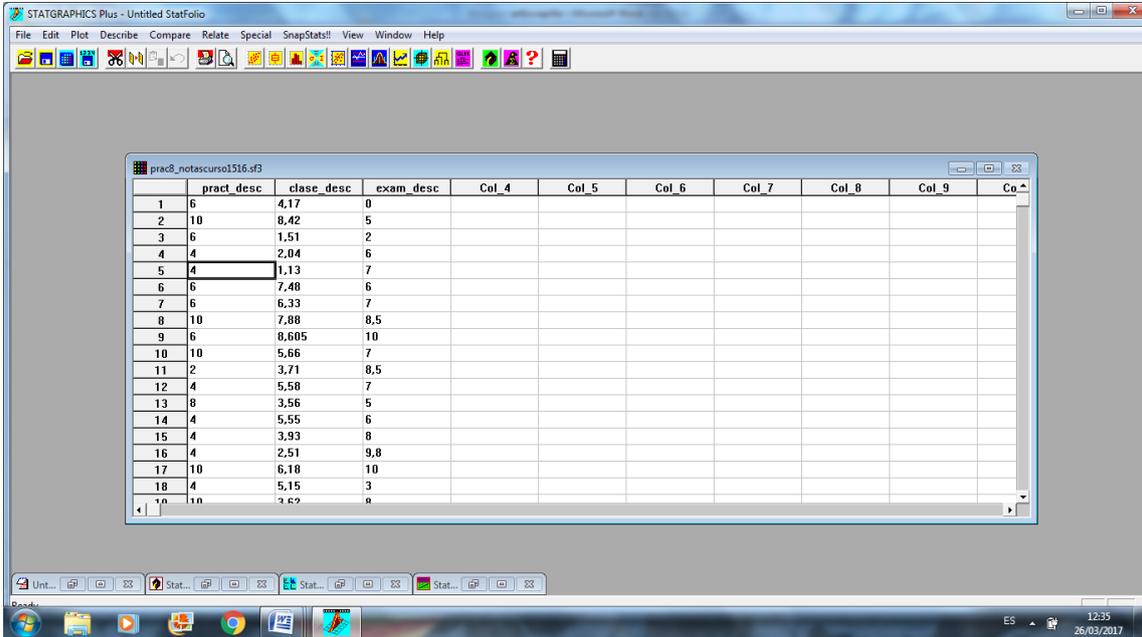
clase_desc: nota del ejercicio sobre estadística descriptiva unidimensional realizado en clase de teoría.

exam_desc: nota de la pregunta del parcial relacionada con estadística descriptiva unidimensional.

Aplicando este procedimiento a las tres columnas del fichero se obtienen los resultados que se detallan en las Figuras 2 a 4. Se observa que:

- En promedio (Average) la mayor nota 6,65 se ha obtenido en la evaluación de la sesión de prácticas de aula informática. La mayor mediana (Median) 7 en la notas de la evaluación del examen.
- La mayor dispersión se ha observado en la evaluación del examen, calculada con la desviación típica 2,74 y el recorrido 10, pero con el intervalo intercuartílico ha sido mayor 6 en la evaluación de prácticas.
- Las tres distribuciones con simétricas, como indican los tres coeficientes de asimetría estandarizados (Stnd. skewness), ya que están en el intervalo (-2, 2).

- En los tres diagramas Box&Whisker se aprecia que no hay datos anómalos.



	pract_desc	clase_desc	exam_desc	Col_4	Col_5	Col_6	Col_7	Col_8	Col_9	Co
1	6	4,17	0							
2	10	8,42	5							
3	6	1,51	2							
4	4	2,04	6							
5	4	1,13	7							
6	6	7,48	6							
7	6	6,33	7							
8	10	7,88	0,5							
9	6	8,605	10							
10	10	5,66	7							
11	2	3,71	8,5							
12	4	5,58	7							
13	8	3,56	5							
14	4	5,55	6							
15	4	3,93	8							
16	4	2,51	9,8							
17	10	6,18	10							
18	4	5,15	3							
19	10	2,69	8							

Figura 1. Fichero con los datos de la muestra

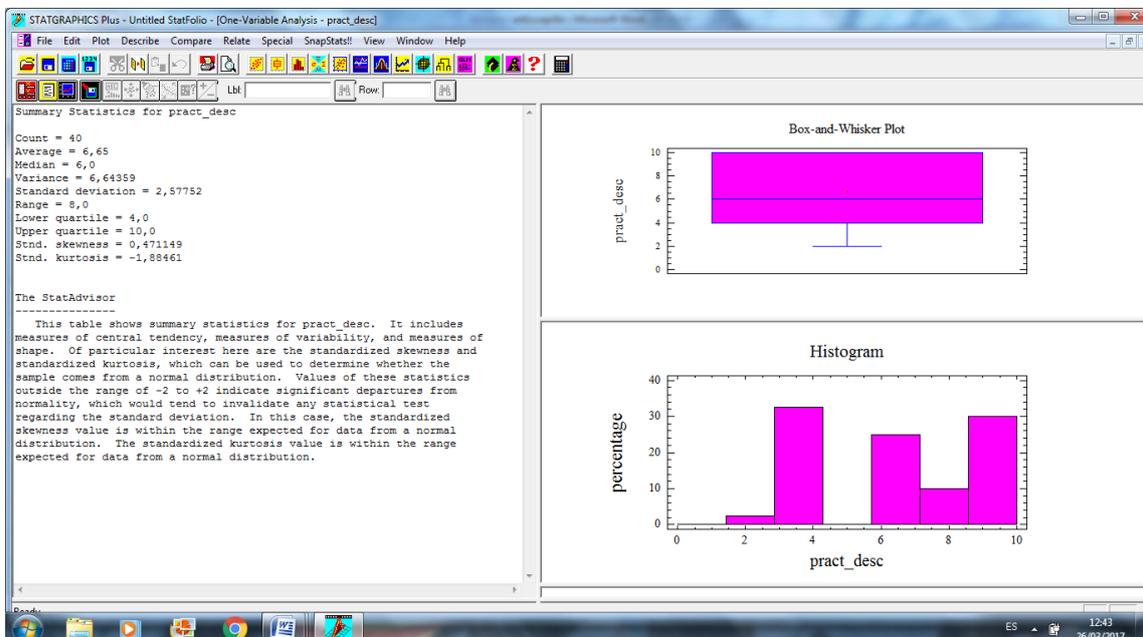


Figura 2. Análisis descriptivo univariante de la evaluación de prácticas

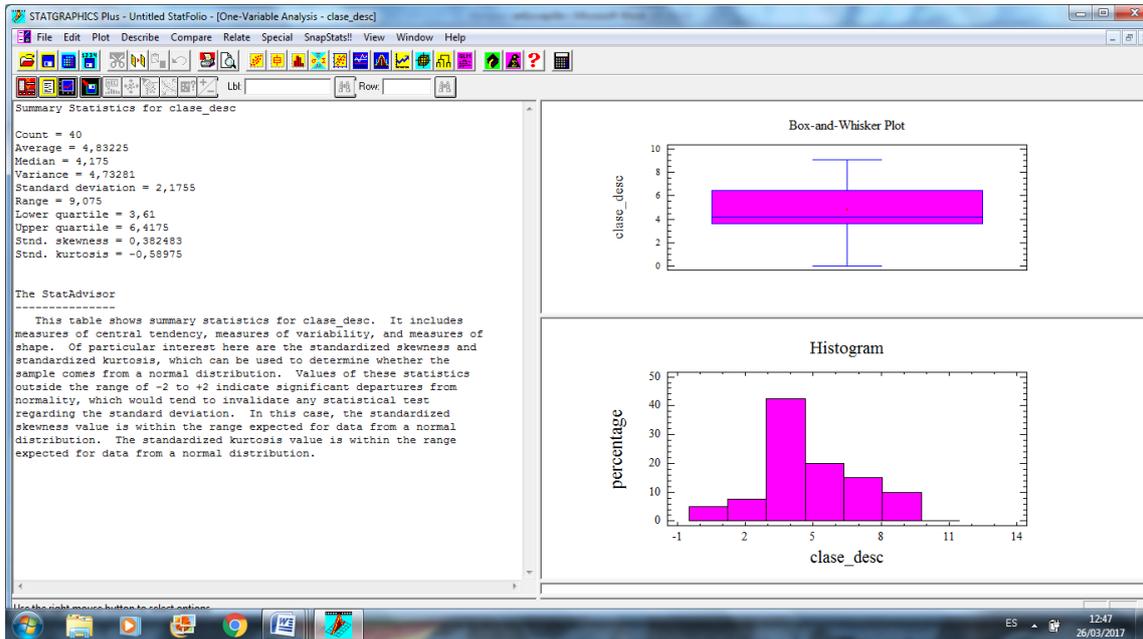


Figura 3. Análisis descriptivo univariante de la evaluación del seminario

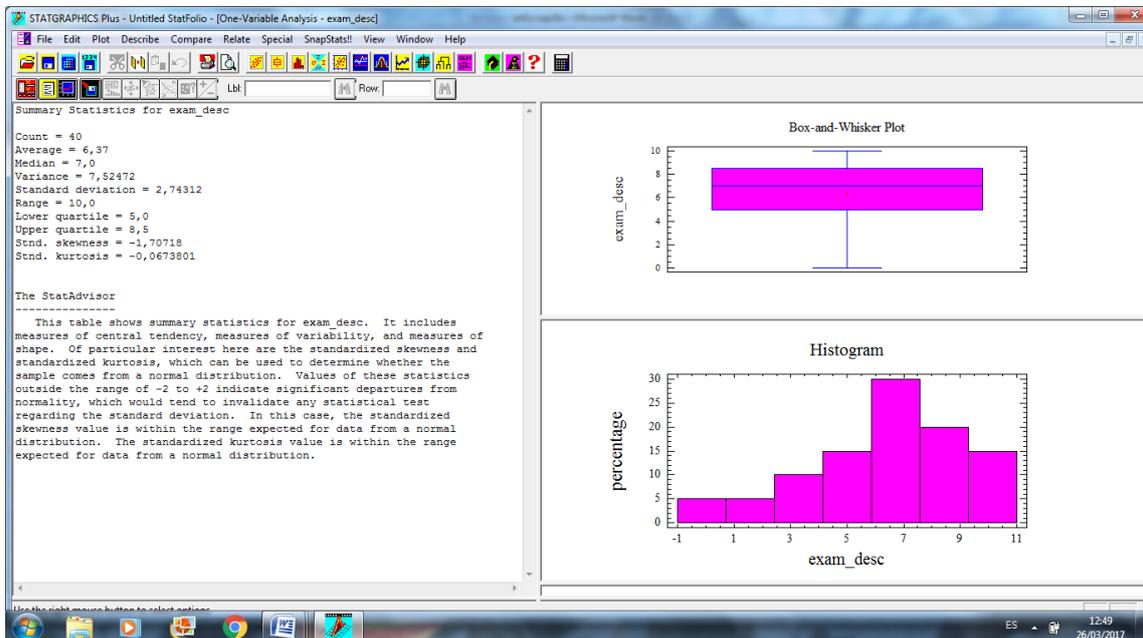


Figura 4. Análisis descriptivo univariante de la evaluación del examen

4.2 Análisis descriptivo bivalente

El análisis descriptivo bivalente permite estudiar la relación que existe entre variables aleatorias. Como en este caso con variables numéricas los métodos aplicar consisten en representar diagramas de dispersión, y calcular la correlación r . Con el programa Statgraphics se obtienen estos resultados en la opción **Describe....Numeric Data....Multiple-Variable Analysis**. La Figura 5 contiene la matriz de diagramas de

dispersión entre los resultados de las evaluaciones. La Figura 6 da la matriz de correlaciones.

A partir de estos resultados se observa:

- Coeficiente de correlación $r=0,3193$ entre la evaluación del seminario (clase_desc) y la del examen (exam_desc) \Rightarrow Hay relación lineal positiva débil, tal y como confirma el diagrama de dispersión.
- Los alumnos con mayor nota en la evaluación de seminario tienen en promedio mayor nota en la pregunta del examen de descriptiva.
- Coeficiente de correlación $r=0,4084$ entre la evaluación de prácticas (pract_desc) y la de seminario (clase_desc) \Rightarrow Hay relación lineal positiva débil. También se observa esto en el gráfico de dispersión. Esta relación es más fuerte que la que existe entre la evaluación del seminario (clase_desc) y la del examen (exam_desc).
- Los alumnos con mayor nota en la evaluación de la práctica de descriptiva tienen en promedio mayor nota en la evaluación de seminario de descriptiva.
- Coeficiente de correlación $r=0,2204$ entre la evaluación de prácticas (pract_desc) y la del examen (exam_desc) \Rightarrow No hay relación lineal. El diagrama de dispersión indica que tampoco hay una relación de otro tipo entre estas evaluaciones.

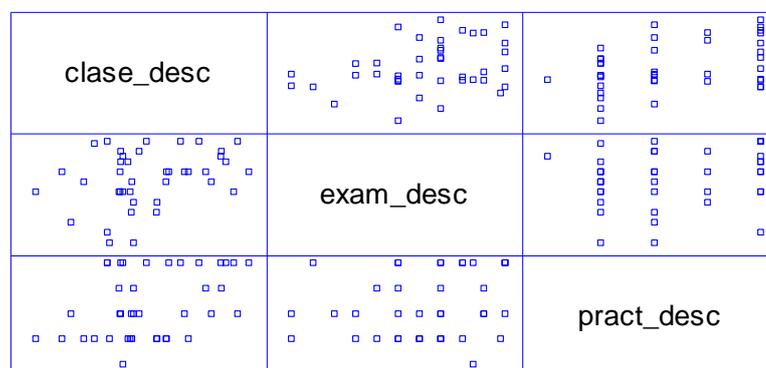


Figura 5. Matriz de diagramas de dispersión

Correlations	clase_desc	exam_desc	pract_desc

clase_desc		0,3193 (40)	0,4084 (40)
		0,0446	0,0089
exam_desc	0,3193 (40)		0,2204 (40)
	0,0446		0,1717
pract_desc	0,4084 (40)	0,2204 (40)	
	0,0089	0,1717	

Correlation (Sample Size)			
P-Value			

Figura 6. Matriz de correlaciones r

4.3 Modelos de regresión lineal simple o rectas de regresión

Para estudiar estas relaciones con más detalle se procede a ajustar una recta de regresión en cada caso, que permita predecir el valor de una de las características en función de otra. Como se hizo en primer lugar la evaluación de seminario, va ser esta la variable explicativa en los dos modelos. Los pasos del análisis son:

- Plantear el modelo de regresión lineal simple.
- Calcular los dos parámetros ordenada y pendiente de la recta. Se utiliza el programa Statgraphics con la opción **Relate....Simple Regression**.
- Estudiar la significación estadística de los dos parámetros utilizando un riesgo de primera especie $\alpha=5\%$.

4.3.1 Recta de regresión que relaciona la nota media de la evaluación del examen con la nota de la de seminario de descriptiva

Planteamiento del modelo:

$$E(\text{exam_desc}/\text{clase_desc}) = a + b \text{ clase_desc}$$

clase_desc: nota del ejercicio sobre estadística descriptiva unidimensional realizado en clase de teoría.

exam_desc: nota de la pregunta del parcial relacionada con estadística descriptiva unidimensional.

$E(\text{exam_desc}/\text{clase_desc})$ = nota media de la evaluación de descriptiva del examen de alumnos que han tenido una nota en la evaluación de seminario de descriptiva igual a clase_desc.

a=ordenada y b=pendiente

Estimación de la recta:

Con la opción de Statgraphics se obtienen las estimaciones de a y b (Figura 8, columna **Estimate**) que indica la recta siguiente

$$E(\text{exam_desc}/\text{clase_desc}) = 4,42 + 0,4 \text{ clase_desc}$$

Estimación de la ordenada $a = 4,42$ = nota media de la pregunta de estadística descriptiva del examen, de alumnos con un 0 en la evaluación de descriptiva del seminario.

Estimación de la pendiente $b = 0,4$ = diferencia en la nota media de la pregunta de estadística descriptiva del examen, de alumnos cuya nota en la evaluación de seminario difiere en un punto.

Regression Analysis - Linear model: $Y = a + b \cdot X$

Dependent variable: exam_desc
Independent variable: clase_desc

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	4,42456	1,02507	4,31635	0,0001
Slope	0,402595	0,193841	2,07693	0,0446

Figura 7. Estimación de la recta

Significación estadística:

El p-valor (columna **P-Value**) de los dos parámetros estimados es menor que el riesgo de primera especie $\alpha=0,05$, por lo que los dos parámetros son significativos estadísticamente. Por tanto, difieren de cero.

4.3.2 Recta de regresión que relaciona la nota media de la evaluación de prácticas con la nota de seminario

Planteamiento del modelo:

$$E(\text{pract_desc}/\text{clase_desc}) = a + b \cdot \text{clase_desc}$$

clase_desc: nota del ejercicio sobre estadística descriptiva unidimensional realizado en clase de teoría.

pract_desc: nota de la evaluación de prácticas relacionada con estadística descriptiva.

$E(\text{pract_desc}/\text{clase_desc})$ = nota media de la evaluación de descriptiva de prácticas de alumnos que han tenido una nota en la evaluación de seminario de descriptiva igual a clase_desc.

a =ordenada y b =pendiente

Estimación de la recta:

Con la opción de Statgraphics se obtienen las estimaciones de a y b (Figura 9, columna **Estimate**) que indica la recta siguiente

$$E(\text{pract_desc}/\text{clase_desc}) = 4,31 + 0,48 \cdot \text{clase_desc}$$

Estimación de la ordenada $a = 4,31$ = nota media de la práctica de estadística descriptiva, de alumnos con un 0 en la evaluación de descriptiva del seminario.

Estimación de la pendiente $b = 0,48$ = diferencia en la nota media de la práctica de estadística descriptiva, de alumnos cuya nota en la evaluación de seminario difiere en un punto.

Significación estadística:

El p-valor (columna **P-Value**) de los dos parámetros estimados es menor que el riesgo de primera especie $\alpha=0,05$, por lo que los dos parámetros son significativos estadísticamente. Por tanto, difieren de cero.



Regression Analysis - Linear model: $Y = a + b \cdot X$

Dependent variable: pract_desc
Independent variable: clase_desc

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	4,31211	0,927783	4,64775	0,0000
Slope	0,483811	0,175444	2,75764	0,0089

Figura 8. Estimación de la recta

5 Cierre

A lo largo de este objeto de aprendizaje hemos visto qué análisis estadísticos se pueden realizar con las evaluaciones de estadística descriptiva. Se han aplicado los métodos descriptivos a su vez, y modelos de regresión lineal simple. Se ha utilizado el programa Statgraphics.

Para comprobar qué realmente has aprendido cómo analizar este tipo de evaluaciones, es el momento de que te pongas manos a la obra e intentes analizar siguiendo los mismos pasos, tus propios resultados. ¡¡ÁNIMO!!

6 Bibliografía

Peña, D: "Estadística modelos y métodos. Vol 1 y 2", Ed. Alianza Universidad, 1995.

Romero Villafranca, R.; Zúñica Ramajo, L: "Métodos estadísticos para ingenieros", Ed. Universidad Politécnica de Valencia, 2013.

Stagraphics Plus 5: "User Manual", Ed. Manugistics Inc., Maryland, USA, 2000.