

Práctica 2

Con mucha frecuencia se ajustan datos mediante modelos de regresión lineal, sin tener en cuenta si se verifican o no las hipótesis del modelo. También, es posible obtener estimaciones muy parecidas para datos peculiarmente distintos. En casos como los mencionados, el **análisis de residuos** resulta fundamental para establecer la validez de los modelos estimados.

En 1973 Anscombe propuso un ejemplo en el que la estimación de diferentes nubes de puntos conduce a la misma recta de regresión; pero el análisis de las gráficas de los valores estimados frente a los predichos, en cada caso, muestra o alguna violación de las hipótesis del modelo, o alguna característica especial que no debe pasarse por alto.

A continuación se recogen los datos de Anscombe:

x	y_1	y_2	y_3	x_d	y_d
10	8,04	9,14	7,46	8	6,58
8	6,95	8,14	6,77	8	5,76
13	7,58	8,74	12,7	8	7,71
9	8,81	8,77	7,11	8	8,84
11	8,33	9,26	7,81	8	8,47
14	9,96	8,1	8,84	8	7,04
6	7,24	6,13	6,08	8	5,25
4	4,26	3,1	5,39	19	12,5
12	10,84	9,13	8,15	8	5,56
7	4,82	7,26	6,42	8	7,91
5	5,68	4,74	5,73	8	6,89

En primer lugar, el alumno creará con Statgraphics el fichero **P2.SF3** con los datos anteriores. Seguidamente comprobará como la regresiones lineales de Y_1 sobre X , Y_2 sobre X , Y_3 sobre X , e Y_d sobre X_d , producen la misma recta

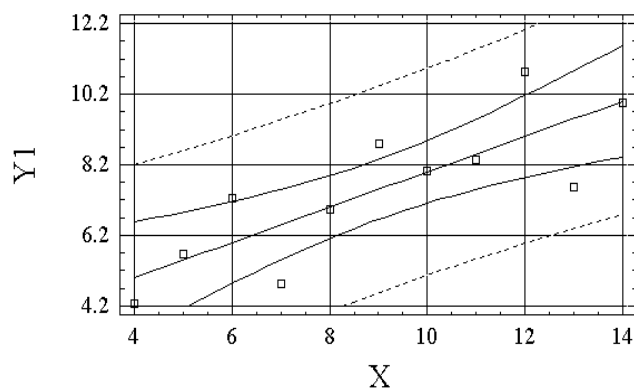
$$\hat{y} = 3 + 0,5x ,$$

con una análisis de la varianza igual en los cuatro casos.

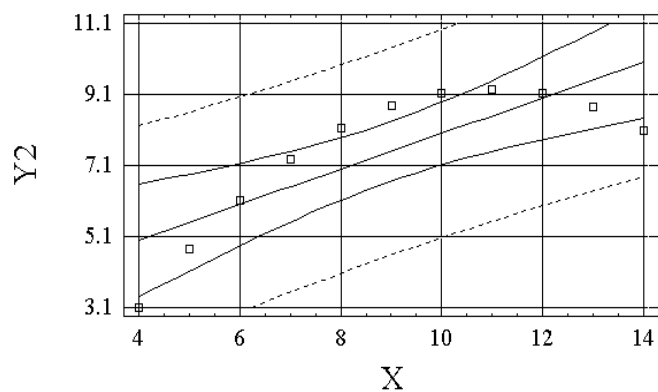
Podemos ver la recta ajustada sobre cada nube de puntos haciendo click sobre el botón **Graphics Options** que aparece debajo de la barra del título de cada ventana **Simple Regression**. Entonces activaremos la casilla de verificación **PLot of Fitted Model**. Las Figuras 1 a 4 muestran los gráficos obtenidos.

Figura 1: Regresión de Y_1 sobre X

Plot of Fitted Model

Figura 2: Regresión de Y_2 sobre X

Plot of Fitted Model

Figura 3: Regresión de Y_3 sobre X

Plot of Fitted Model

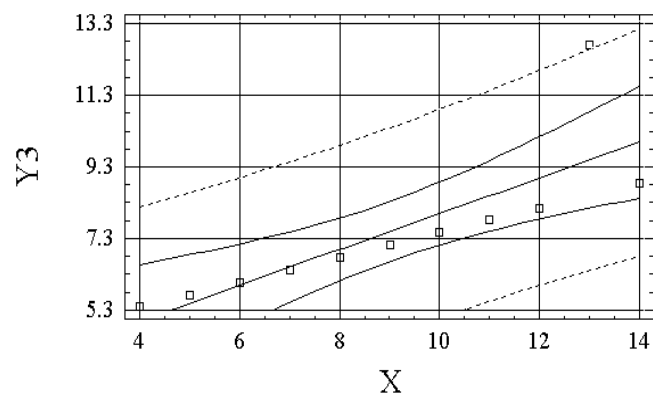
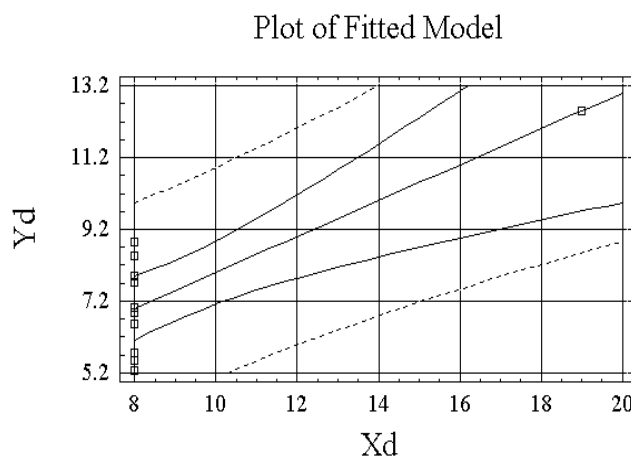


Figura 4: Regresión de Y_d sobre X_d 

En la situación anterior se podría pensar que las cuatro regresiones son idénticas. Vamos a estudiar los residuos de cada regresión, lo que nos permitirá comprobar que cada caso es bien distinto. Para ello obtendremos las gráficas de los residuos frente a la variable independiente, y de los residuos frente a los valores predichos. Tales gráficas, que se recogen en las Figuras 5 a 12, se obtienen haciendo click sobre el botón **Graphics Options** y activando las casillas de verificación **Residuals versus X** y **Residuals versus Predicted**.

La observación de dichas gráficas permite extraer ciertas conclusiones:

- El primer modelo no ofrece ninguna clara evidencia de anomalías importantes.
- El segundo modelo refleja una falta clara de linealidad pues se observa una tendencia curvilínea en los residuos.
- El tercer modelo presenta una observación atípica que seguramente determina la estimación de la recta. En concreto, afecta claramente a la pendiente de regresión.
- El cuarto modelo presenta también un valor atípico pero, a diferencia del tercer modelo, se puede observar que la pendiente de la recta viene determinada únicamente por un punto, lo cual muestra muy poca fiabilidad en el empleo de dicha recta con fines predictivos. Además, resulta evidente que dicho punto determina la propia existencia de la recta de regresión, a diferencia de lo que ocurría en el tercer modelo.

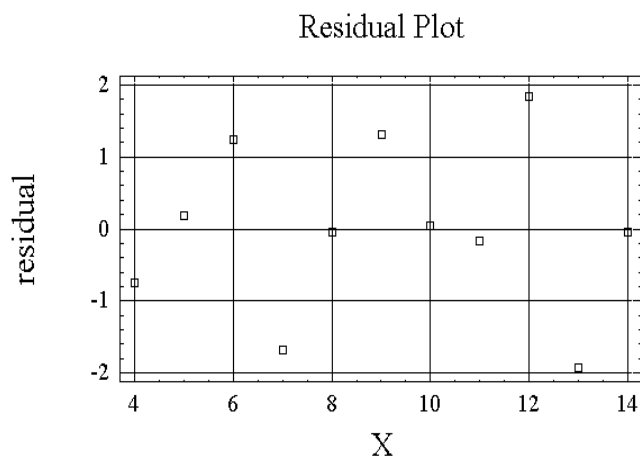
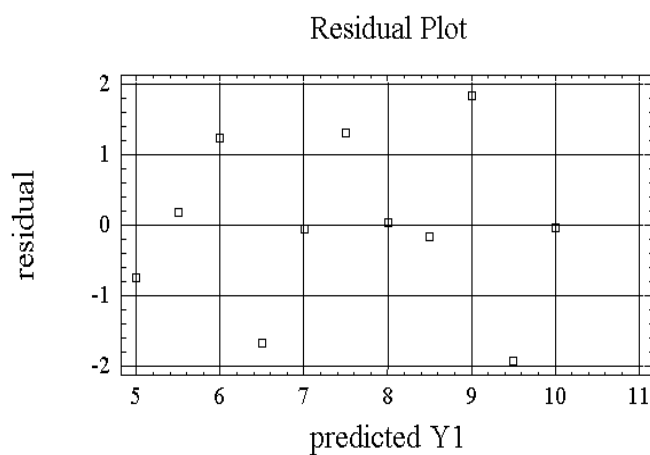
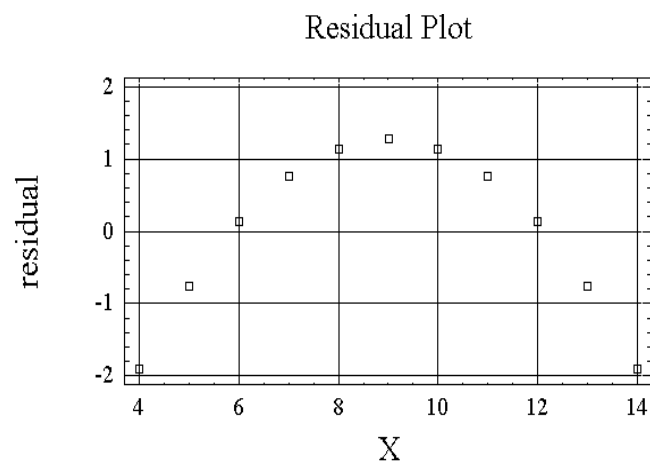
Figura 5: Residuos frente a la variable explicativa en la regresión de Y_1 sobre X Figura 6: Residuos frente a los valores predichos en la regresión de Y_1 sobre X Figura 7: Residuos frente a la variable explicativa en la regresión de Y_2 sobre X 

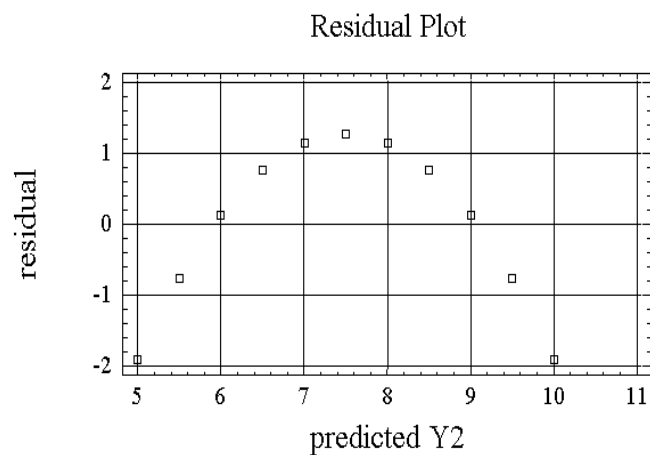
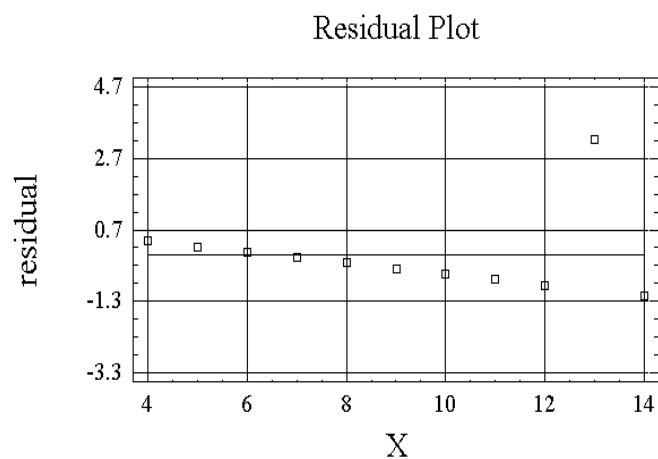
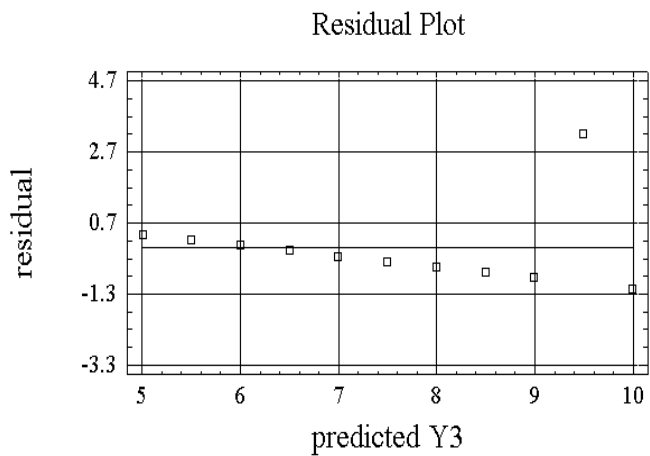
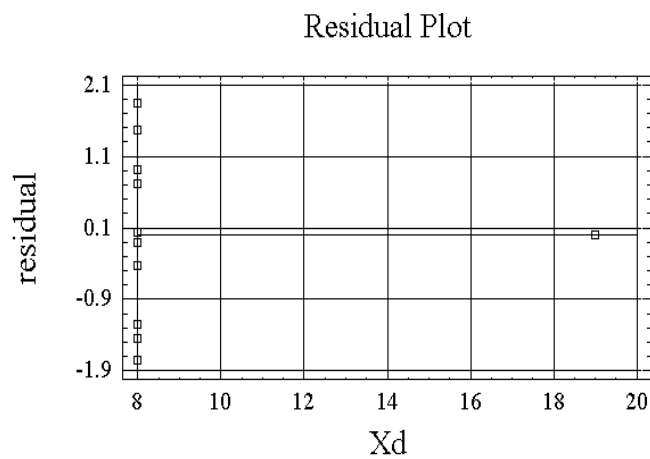
Figura 8: Residuos frente a los valores predichos en la regresión de Y_2 sobre X Figura 9: Residuos frente a la variable explicativa en la regresión de Y_3 sobre X Figura 10: Residuos frente a los valores predichos en la regresión de Y_3 sobre X 

Figura 11: Residuos frente a la variable explicativa en la regresión de Y_d sobre X_d Figura 12: Residuos frente a los valores predichos en la regresión de Y_d sobre X_d 