

Ejemplo de regresiones.

En el archivo *ejemplo_regresion.xls*, en la hoja "Mediciones" se muestran los resultados de 3 mediciones distintas de la variable dependiente y en función de la variable independiente x . Las mediciones distintas se refieren al posible uso de instrumentos y métodos de medición distintos.

Cada una de las mediciones se ajustan con 3 modelos distintos que son los siguientes

Modelo 1, lineal

$$y = mx + b$$

Modelo 2, seno

$$y = A \sin\left(\pi \frac{x - xc}{w}\right)$$

Modelo 3, seno más constante

$$y = B + A \sin\left(\pi \frac{x - xc}{w}\right)$$

Usando el Origin se pueden hacer las regresiones para las 3 mediciones y los 3 modelos (un total de 9 ajustes). Entre las opciones de los datos de salidas, además de los parámetros del ajuste, se puede visualizar los valores del F -test así como los de los t -test para cada uno de los parámetros y el análisis del residuo.

Si bien el Origin (o el programa que se utilice para la regresión) muestra los resultados finales, a fin de visualizar mejor el significado del proceso de ajuste, como ejemplo, se usa la medición 1 con el modelo 1 (hoja "Medicion1_Modelo1") donde se muestra el cálculo de TSS , RSS , ESS , R^2 , F y t junto a los grados de libertad, los criterios basados en los valores de p , y el residuo (que también se muestra en el archivo *medicion1_modelo1.png*). Como puede verse, el modelo lineal ajusta muy bien en este caso, y los valores de p son mucho menores que 0,05. El test de normalidad de residuos puede no ser muy útil en este caso con tan pocos puntos experimentales, en cuyo caso el criterio visual es más práctico. Debe entenderse que el análisis de residuos está pensado para detectar anomalías más que para confirmar casos normales.

Distinto es el caso de la medición 2 con el modelo 2 (hoja "Medicion2_Modelo2"), donde si bien se determina que hay una relación entre x e y (porque el valor de p del F -test es menor que 0,05), existe un parámetro (xc) del ajuste de la función seno que no es estadísticamente significativo (el valor de p del t -test es mayor que 0,05). Además, el residuo (que también se muestra en el archivo *medicion2_modelo2.png*) no parece estar distribuido normalmente, lo cual indica que el modelo no es adecuado, o que la incerteza de los valores medidos no sigue una distribución normal. Una indicación de que esto último ocurre, lo da el hecho que se empieza a ver que $TSS \neq ESS + RSS$, cuando se esperaría $TSS = ESS + RSS$. En regresiones lineales esta última es una igualdad matemática, pero en un caso general nunca se cumple exactamente porque una muestra con una cantidad finita de puntos no está perfectamente distribuida según una gaussiana, pero, cuanto más diferencia haya, más los datos se apartan de una distribución gaussiana.

El análisis de estos 2 casos más los 7 restantes (que quedan a cargo del lector) permite extraer la información relevante de las regresiones y los residuos que llevan a entender tanto aspectos del modelo como de los datos experimentales. ¿Cuáles serían las principales conclusiones de las 3 mediciones y los 3 modelos?