

Favor de abrir el navegador Mozilla Firefox y escriba la siguiente dirección <http://math.uprag.edu/area.MTW> para obtener el conjunto de datos.

Vamos a utilizar la opción **Fitted line Plot** del menú de **Regression** para hallar bandas de confianza tanto para el valor predicho como para el valor medio de las Y. Para esto se deben elegir las opciones **Display Confidence Interval** y **Display Prediction Interval** al oprimir el boton **Options**. Con las bandas de confianza se pueden tener intervalos de confianzas para cualquier valor dado de X. Para el presente ejemplo se obtiene:

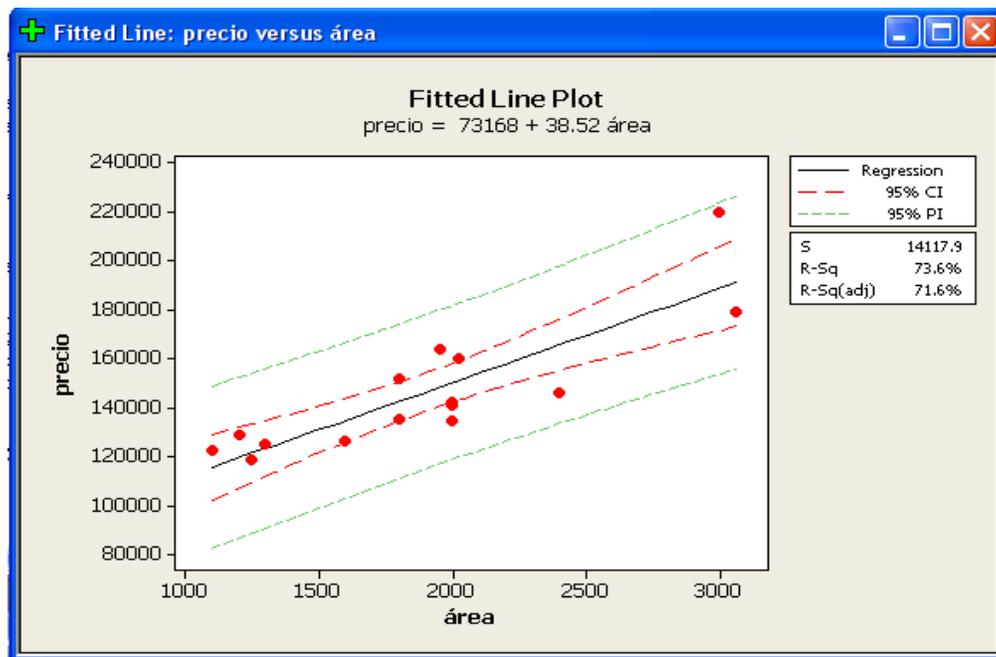


Figura 9.5 Bandas de Confianza para el valor medio y el valor predicho de Y

Notar que las bandas de confianza son anchas en los extremos del eje X y angostas en el centro del mismo. En realidad las bandas se van angostando cuando los valores de X que se toman están cerca del promedio \bar{x} .

9.3 Análisis de Residuales

Un residual r_i es la diferencia entre el valor observado Y_i y el valor estimado por la línea de regresión \hat{Y}_i , es decir, $r_i = Y_i - \hat{Y}_i$. El residual puede ser considerado como el error aleatorio e_i observado. También se acostumbra usar el **Residual estandarizado**, el cual se obtiene al dividir el

residual entre la desviación estándar del residual (siempre que hagamos análisis de residuales debemos utilizar **Residual estandarizado**), y el **Residual estudentizado "deleted"**, que es similar al anterior pero eliminando de los cálculos la observación cuyo residual se desea hallar.

El análisis de residuales permite cotejar si las suposiciones del modelo de regresión se cumplen.

Se puede detectar:

- a) Si efectivamente la relación entre las variables X e Y es lineal.
- b) Si hay normalidad de los errores.
- c) Si hay valores anormales en la distribución de errores (Si se usa **Residual estandarizado**, cualquier observación con un residual mayor de 2 o menor de -2 es considerado “**outlier**”)
- d) Si hay varianza constante (propiedad de Homocedasticidad) y
- e) Si hay independencia de los errores.

El análisis de residuales se puede llevar a cabo gráficamente o en forma analítica. En este texto sólo consideraremos un análisis gráfico, las cuales pueden obtenerse de dos maneras. La primera manera es escogiendo el botón **Graphs** de la ventana de diálogo **Regression**.

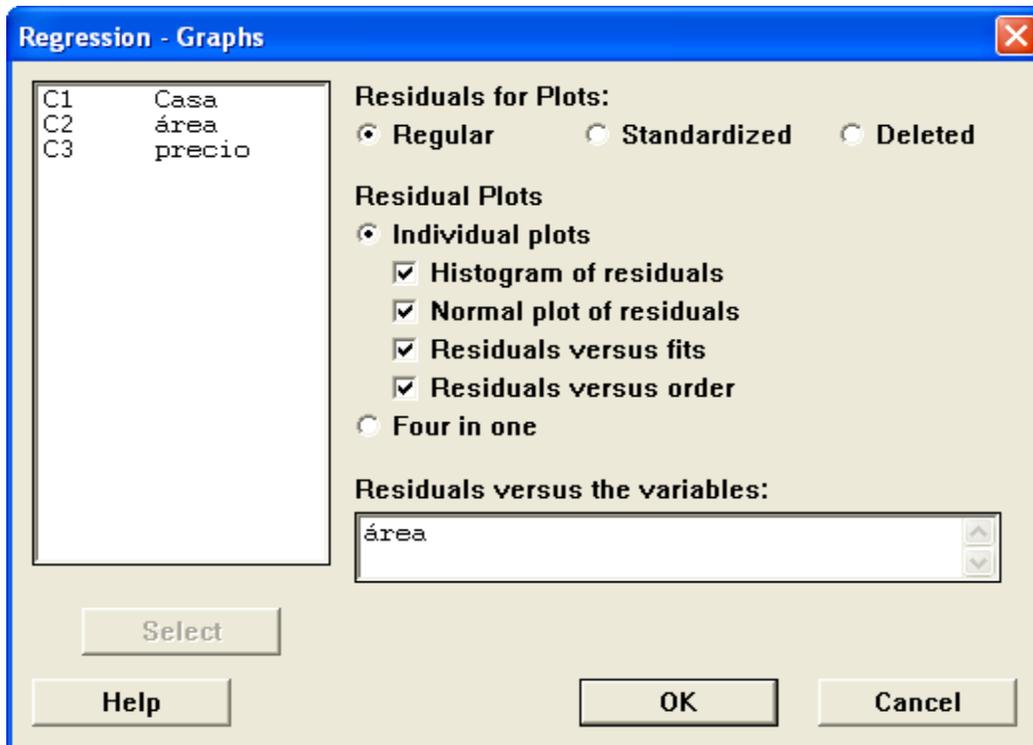


Figura 9.6. Ventana de diálogo que aparece al oprimir **Graphs** en **Regression**.

Hay tres posibles elecciones de residuales y hasta 5 plots de residuales que se pueden hacer. Las ventanas de gráficas aparecerán en cascada como se muestra en la Figura 9.7. En esta manera cada plot de residual sale en una ventana gráfica separada. Hay cinco plots que se usan:

- 1) **Plot de Normalidad:** Permite cotejar normalidad. Si los puntos están bien cerca de una línea recta

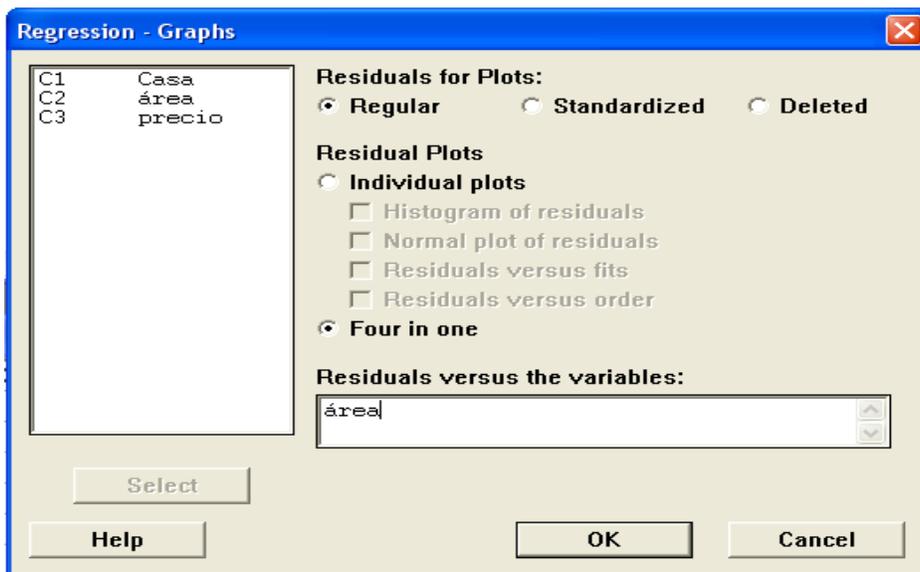


Figura 9.8. Ventana de diálogo de la opción **graph** de **regression**.

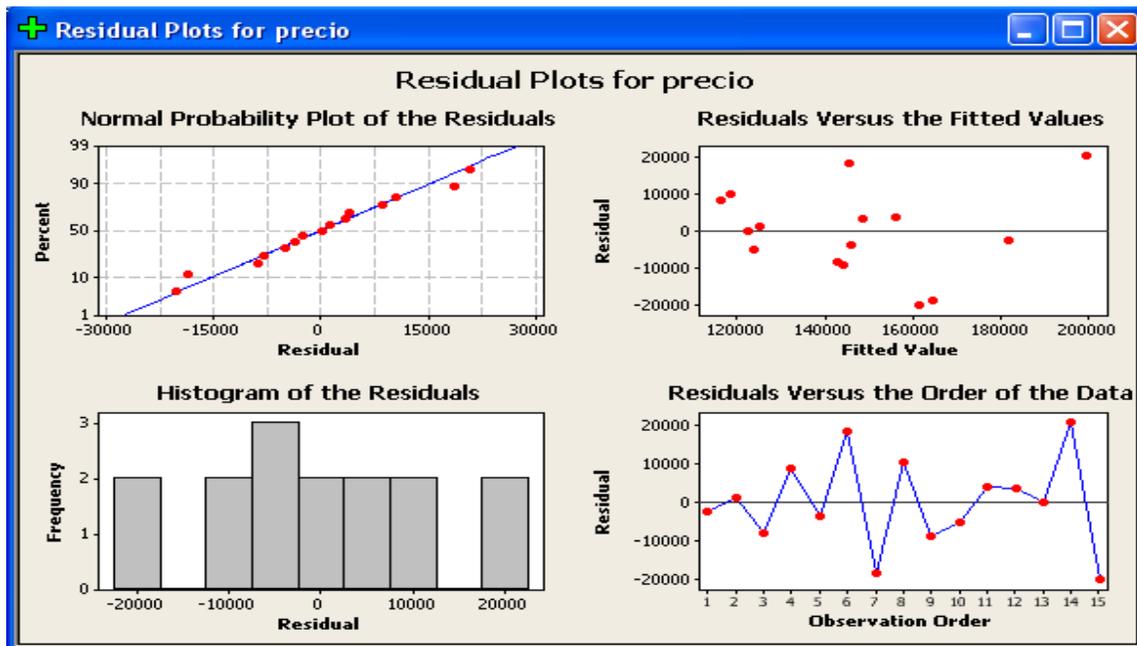


Figura 9.9. Plots de residuales en una misma ventana.

Aparecerán en una misma página los cuatro primeros plots de la lista mencionada anteriormente, como se muestra en la Figura 9.9.

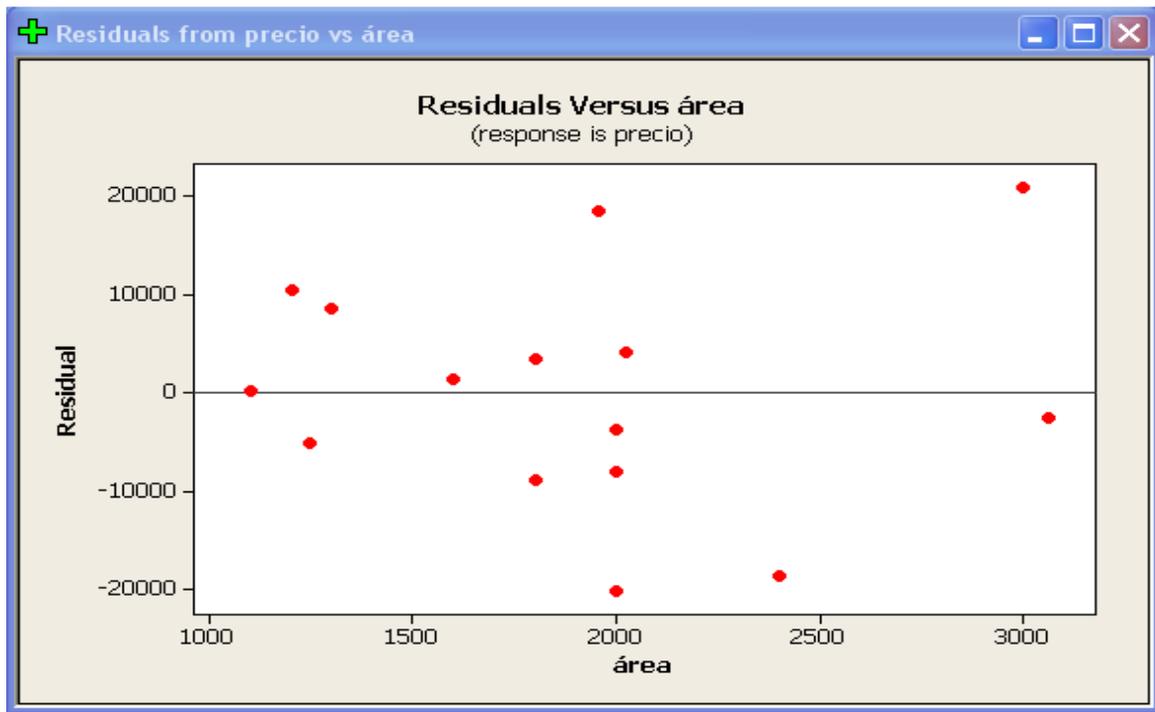


Figura 9.10. Plots de residuales versus la variable area.

Interpretación: Los puntos del plot de normalidad no caen cerca de una línea recta y en el extremo superior se detecta un “outlier”. Similarmente, el histograma no es simétrico con un pico central y también muestra un “outlier” en el extremo superior. En conclusión, no hay normalidad de los errores. El plot de residuales versus el índice de la observación muestra que la observación 14 es un “outlier”, pues el residual estandarizado cae más allá de dos. El plot de los residuales versus los valores predichos muestra que la varianza de los errores no es constante con respecto a la variable de respuesta, pues tiende a aumentar cuando el valor de la variable de respuesta aumenta.

Hay maneras de corregir algunas de las anomalías encontradas en el análisis de residuales, las cuales pueden ser leídas en un texto especializado de regresión.