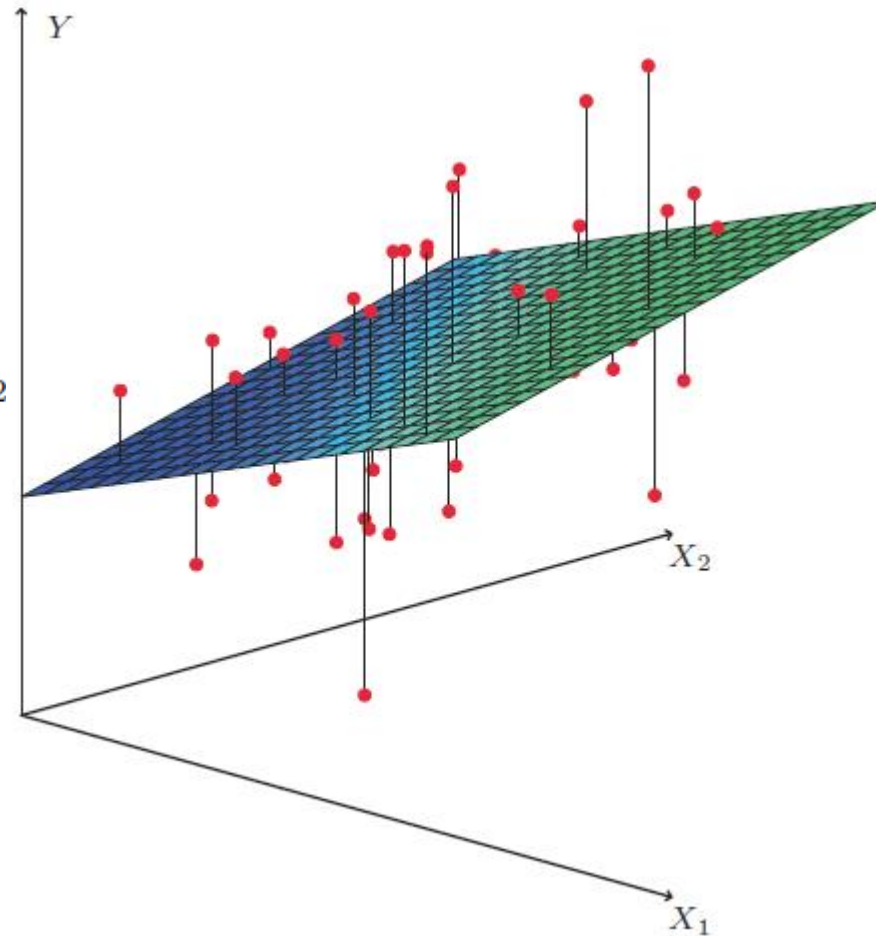


$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

$$\begin{aligned} \text{RSS} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2 \end{aligned}$$

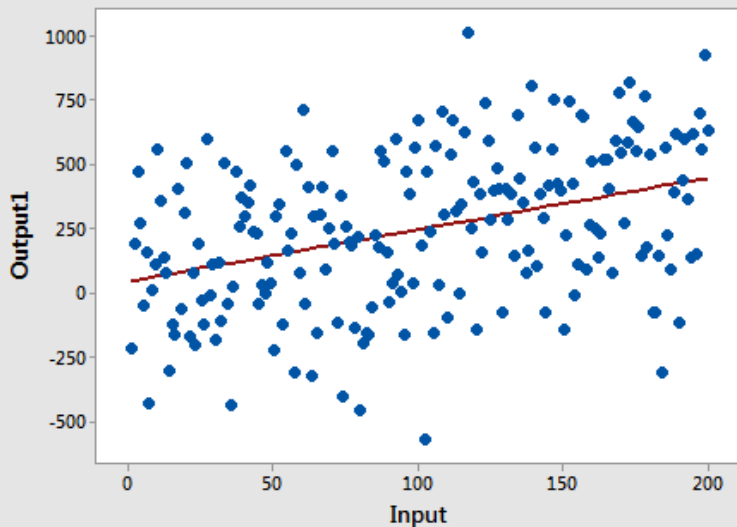


¿Los valores bajos de R^2 son intrínsecamente malos?

Hay dos razones principales por las que puede estar bien tener valores bajos de R^2 . En algunos campos, se espera que los valores de R^2 sean bajos. Por ejemplo, en psicología, típicamente se obtienen valores de R^2 inferiores al 50%. No así en tecnología e industria

Comparación de modelos de regresión con valores de R^2 bajos y altos

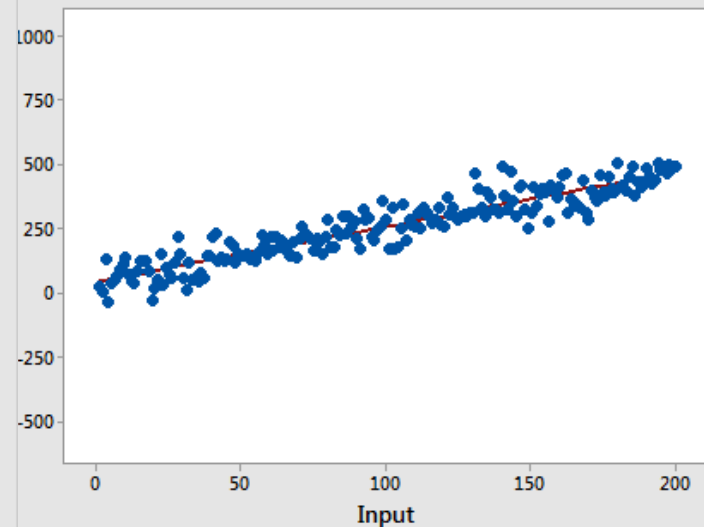
Fitted Line Plot
Output1 = 44.53 + 2.024 Input



S	283.042
R-Sq	14.7%
R-Sq(adj)	14.3%



Fitted Line Plot
Output2 = 44.86 + 2.134 Input

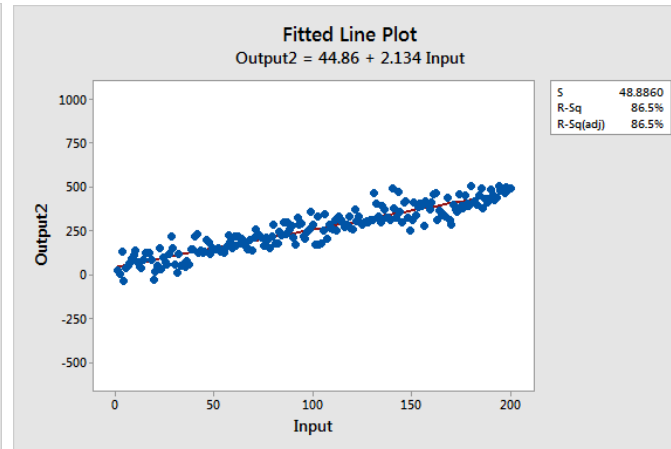
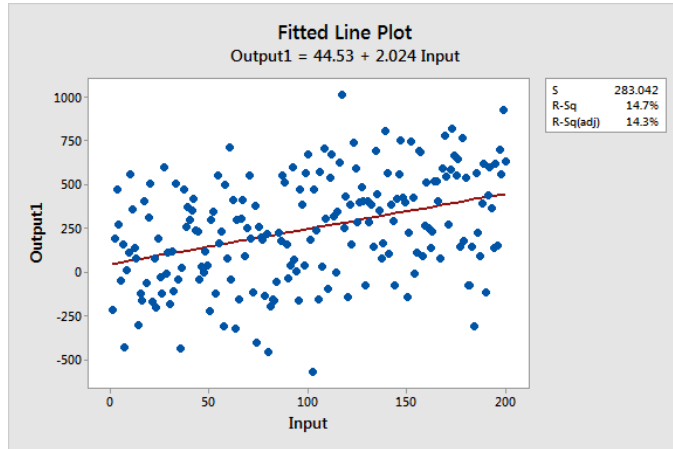


S	48.8860
R-Sq	86.5%
R-Sq(adj)	86.5%

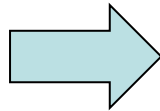


Prueba de Hipótesis sobre los Coeficientes

El error estándar de los residuos(RSE) mide la distancia media entre los valores observados y la línea de regresión.

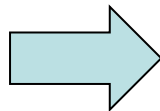


Error standard (s)
de los coeficientes



$$SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad SE(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Estimo el error
standard de la
regresión con el RSE



$$RSE = \sqrt{RSS/(n - 2)}$$

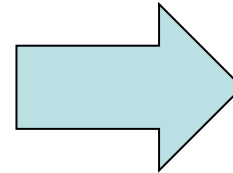


Los errores estándar pueden utilizarse para realizar pruebas de hipótesis sobre los coeficientes

H_0 : No hay relación entre X e Y

H_a : Hay alguna relación entre X e Y

Matemáticamente, esto corresponde a la prueba

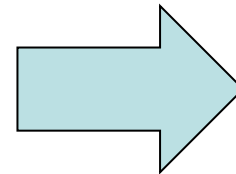


$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Si $\beta_1 = 0$ entonces el modelo se reduce a $Y = \beta_0 + \varepsilon$ (Y no depende de X). Para probar la hipótesis nula necesitamos determinar si $\hat{\beta}_1$, nuestra estimación para β_1 , está lo suficientemente lejos de cero como para que podamos con fiar en que β_1 no es cero. ¿Cuán lejos es suficiente? Esto, por supuesto depende de la exactitud de $\hat{\beta}_1$, es decir, depende de $SE(\hat{\beta}_1)$.

En la práctica, calculamos un estadístico t



$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$



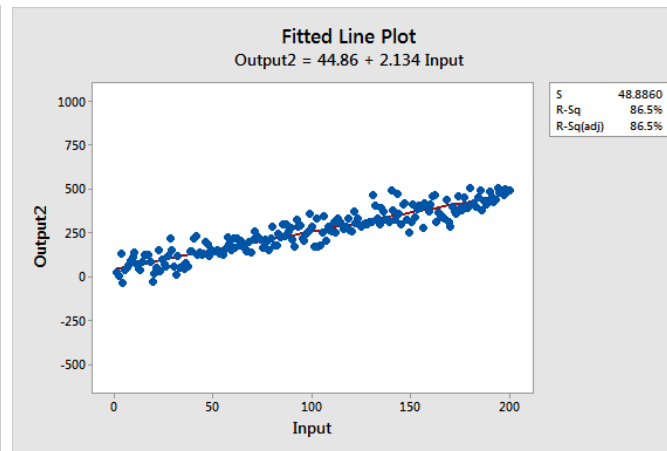
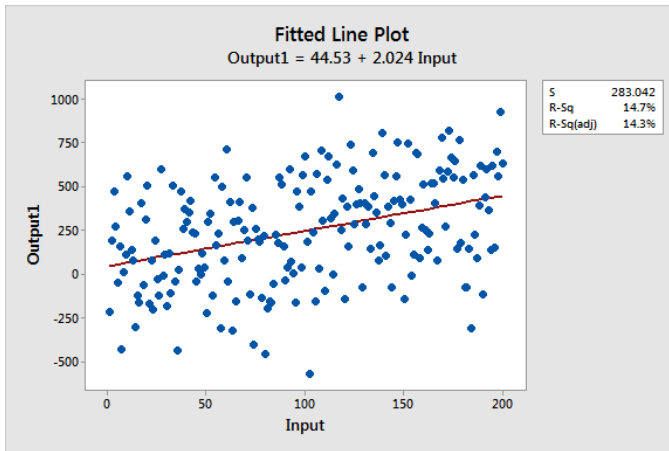
¿Cómo interpreto los *valores-p* en el análisis de regresión lineal?

El *valor-p* para cada término prueba la hipótesis nula de que el coeficiente es igual a cero (sin efecto). Un *valor-p* bajo ($< 0,05$) indica que puede rechazar la hipótesis nula. En otras palabras, un predictor que tiene un *valor-p* bajo es probable que sea una adición significativa a su modelo porque cambios en su valor se relacionarán con cambios en la variable de respuesta. Por el contrario, un *valor-p* mayor (insignificante) sugiere que los cambios en el predictor no están asociados con cambios en la respuesta.

Coefficients

Term	Coef	SE Coef	T	P
Constant	389.166	66.0937	5.8881	0.000
East	2.125	1.2145	1.7495	0.092
South	5.318	0.9629	5.5232	0.000
North	-24.132	1.8685	-12.9153	0.000





Los dos modelos son casi idénticos en varios aspectos:

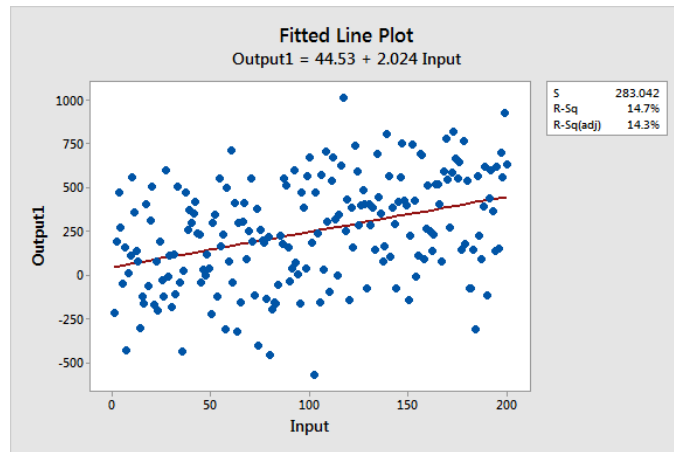
- Ecuaciones de regresión:
 $salida = 44 + 2 \cdot entrada$
- La variable *entrada* es significativa con $P < 0.001$ para ambos modelos

La interpretación del *valor-p* y del coeficiente $\hat{\beta}_1$ (asociado a la variable *entrada*) no cambian. Si se mueve hacia la derecha en cualquiera de las dos líneas aumentando *entrada* en una unidad, se produce un aumento medio de dos unidades en *salida*. Para ambos modelos, el valor significativo de *p* indica que se puede rechazar la hipótesis nula de que el coeficiente es igual a cero (sin efecto).

Además, si ingresamos el mismo valor de *entrada* en ambas ecuaciones, la regresión devolverá predicciones casi equivalentes para *salida*. Por ejemplo, una *entrada* de 10 produce una predicción de *salida* de 66,2 para un modelo y de 64,8 para el otro.



Para evaluar la precisión deberemos calcular los intervalos de **predicción**. Un intervalo de predicción es un rango que probablemente contenga el valor de respuesta de una **nueva observación** dada la configuración especificada de los predictores en el modelo. A continuación se muestran los valores ajustados y los intervalos de predicción para una *entrada* de 10.



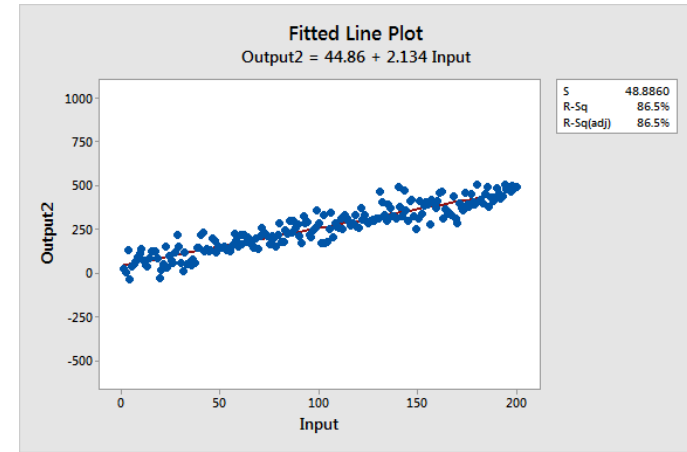
Prediction for Output1

Regression Equation

$$\text{Output1} = 44.5 + 2.024 \text{ Input}$$

Variable	Setting
Input	10

Fit	SE Fit	95% CI	95% PI
64.7766	37.2129	(-8.60793, 138.161)	(-498.190, 627.743)



Prediction for Output2

Regression Equation

$$\text{Output2} = 44.86 + 2.1343 \text{ Input}$$

Variable	Setting
Input	10

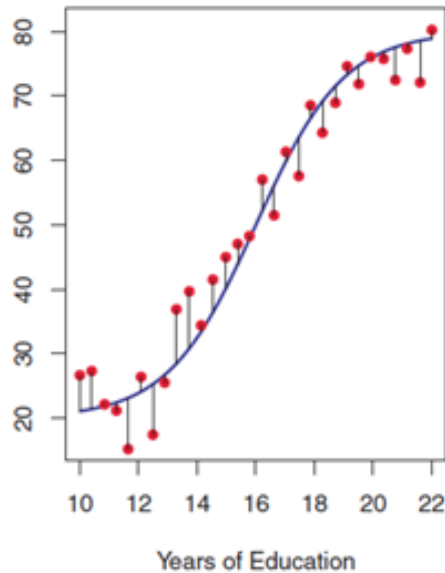
Fit	SE Fit	95% CI	95% PI
66.2076	6.42728	(53.5329, 78.8823)	(-31.0260, 163.441)



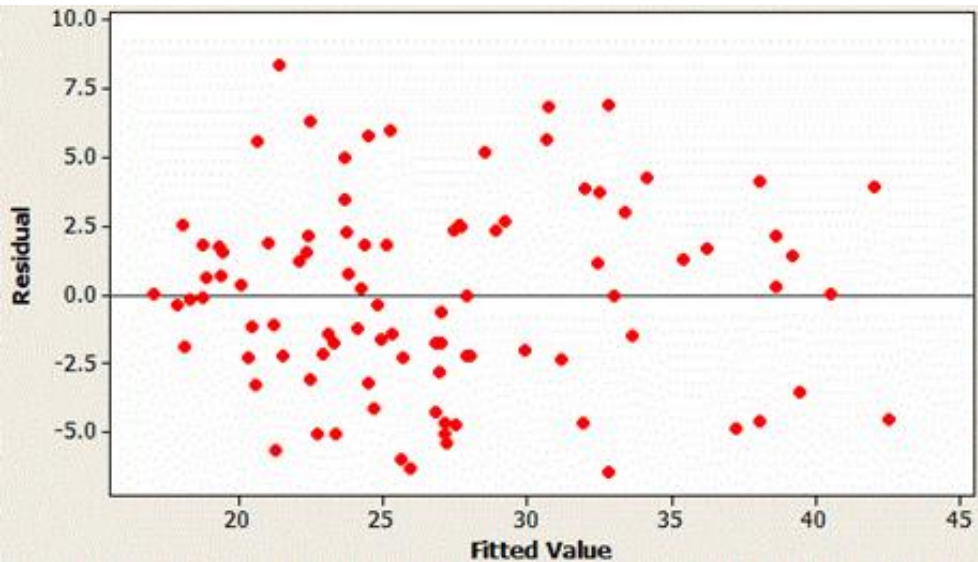
La regresión lineal calcula una ecuación que minimiza la distancia entre la línea ajustada y todos los puntos de datos. Técnicamente, la regresión de mínimos cuadrados ordinarios (OLS) minimiza la suma de los residuos cuadrados.

En general, un modelo se ajusta bien a los datos si las diferencias entre los valores observados y los valores pronosticados del modelo son pequeñas e **insesgadas**

Gráficos de Residuos

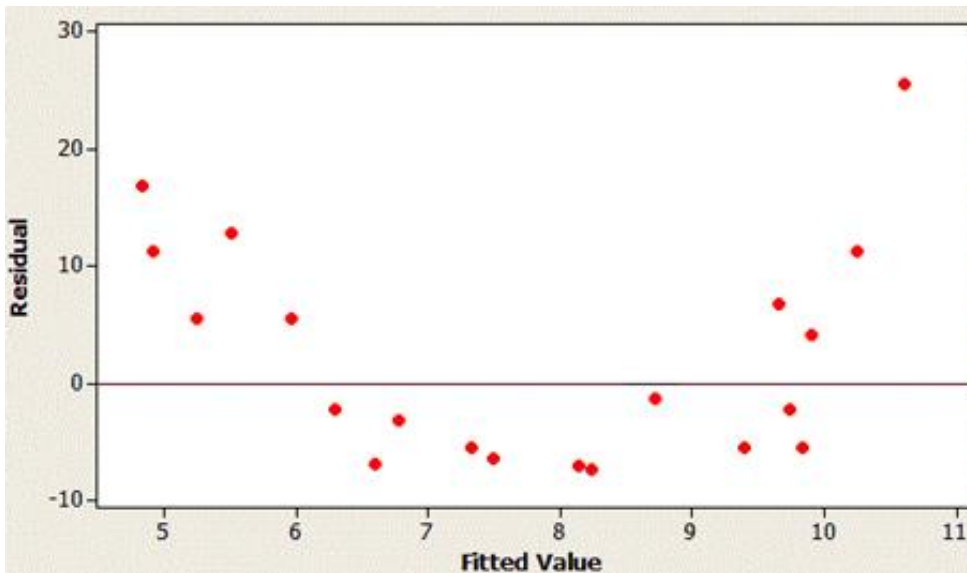


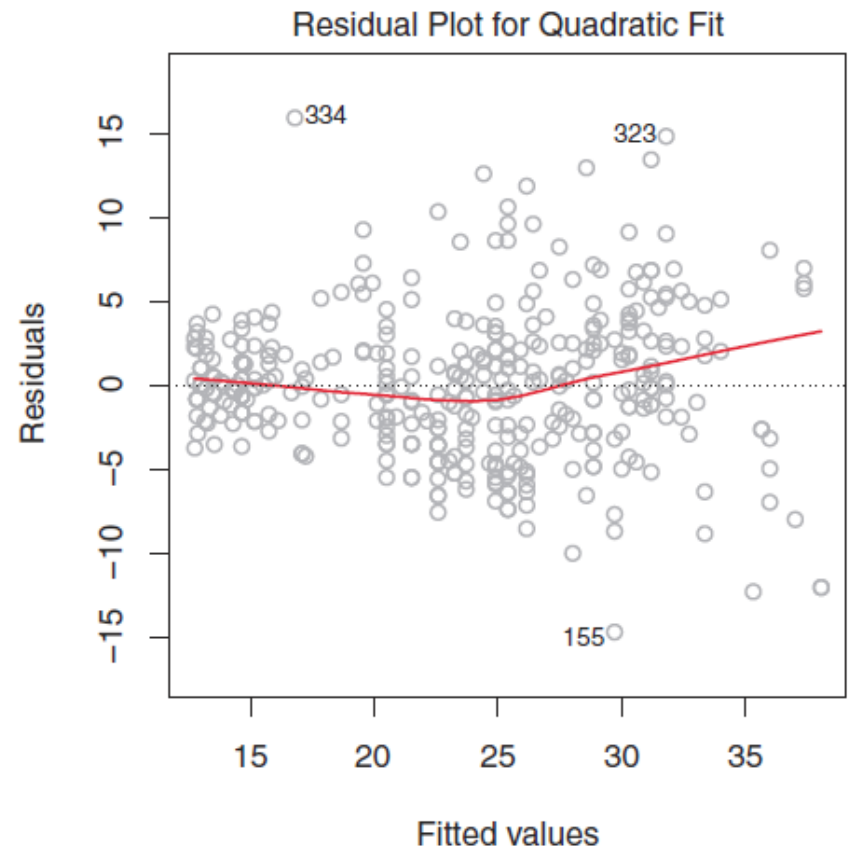
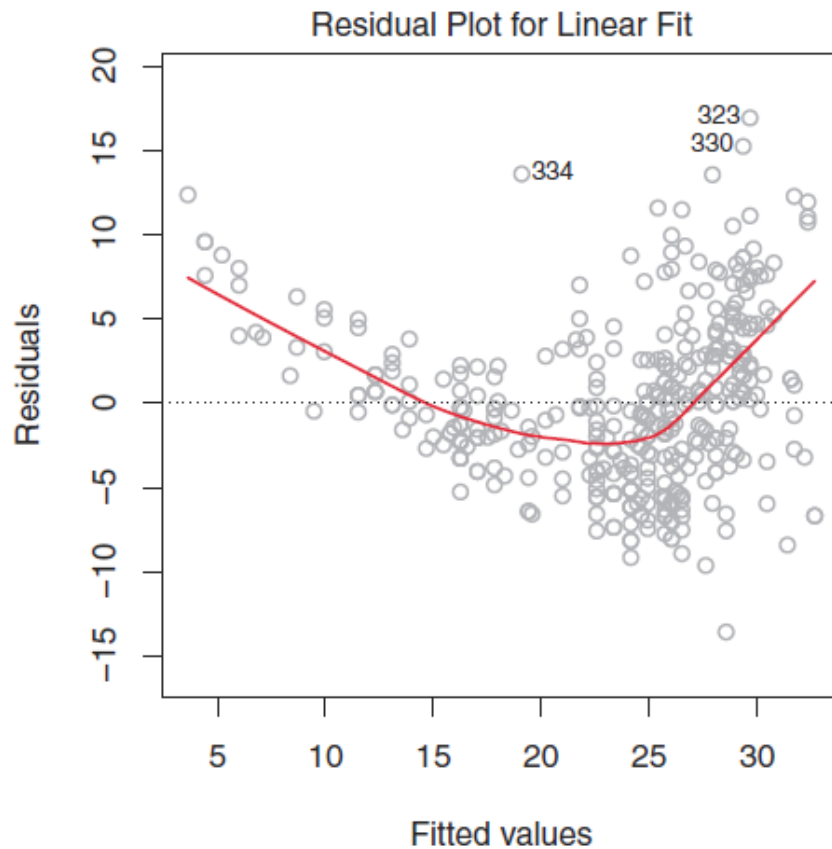
El error es la diferencia entre el valor esperado y el valor observado. En conjunto, las diferencias entre los valores esperados y los observados deben ser impredecibles. Nada de la información explicativa/predictiva (ni el modelo real ni nuestra estimación paramétrica de este) debería estar en el error.



El patrón no aleatorio en los residuos indica que la porción determinista (variables predictoras) del modelo no está capturando alguna información explicativa que se está "filtrando" a los residuos. El gráfico podría representar varias maneras en las que el modelo no está explicando todo lo que es posible. Las posibilidades incluyen:

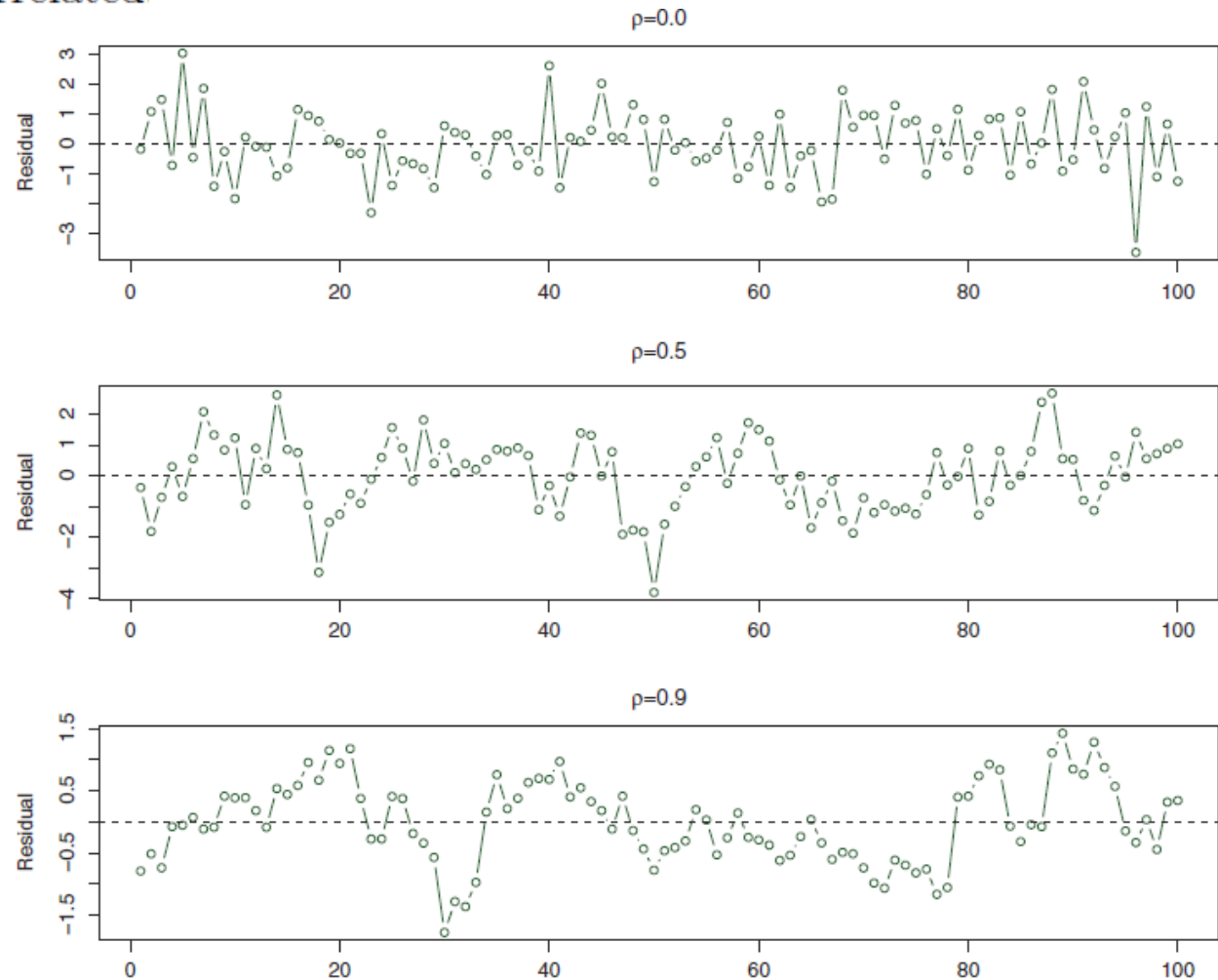
- Falta de un término de orden superior de una variable en el modelo para explicar la curvatura.
- falta de interacción entre los términos que ya están en el modelo



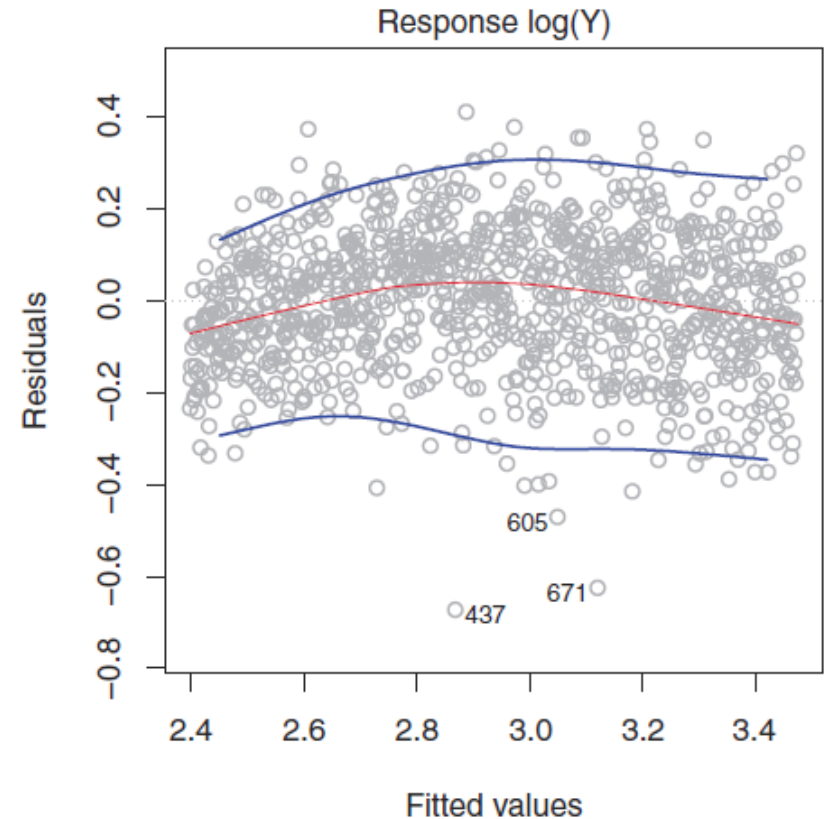
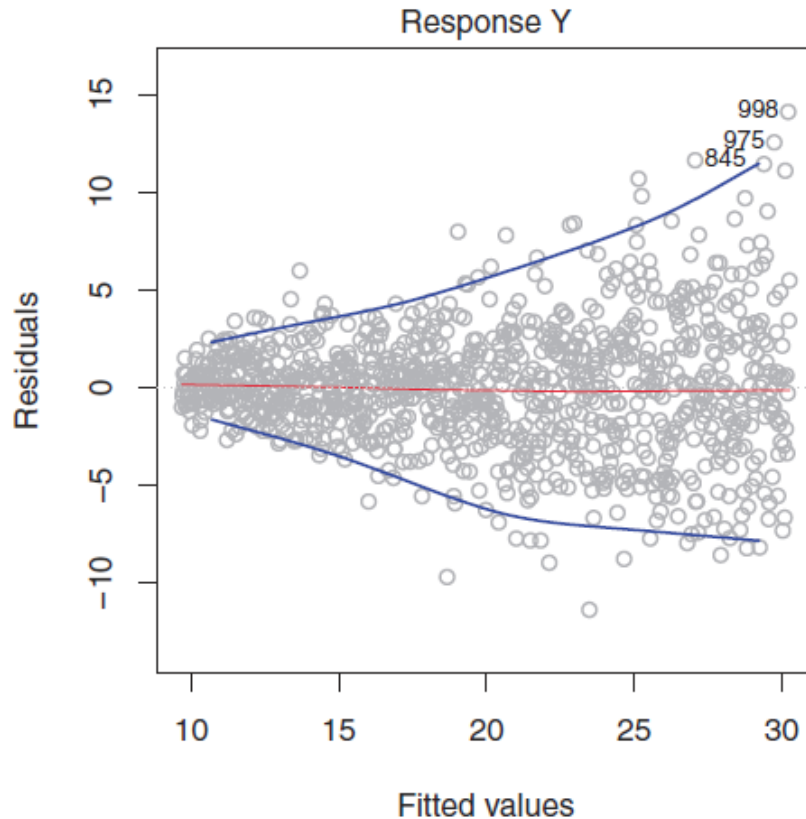


Correlación entre los residuos

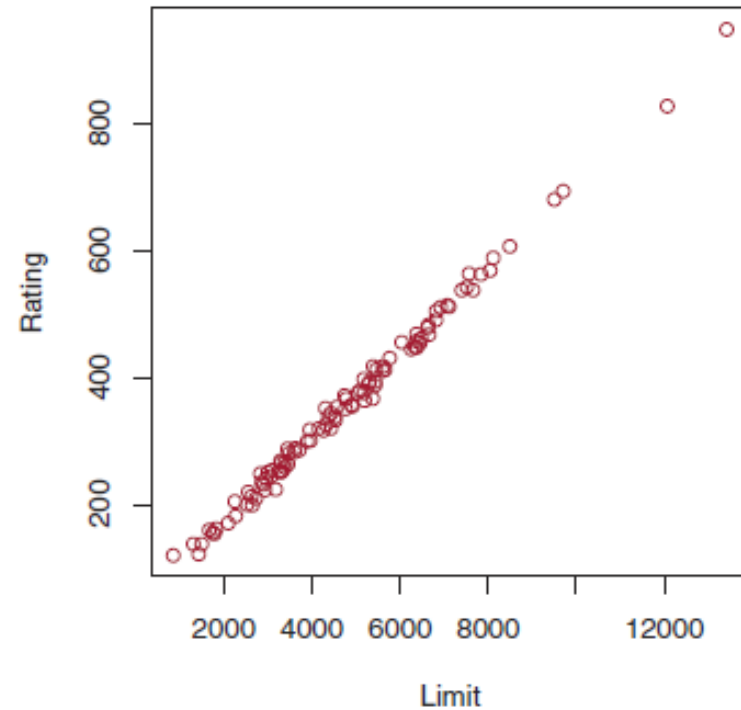
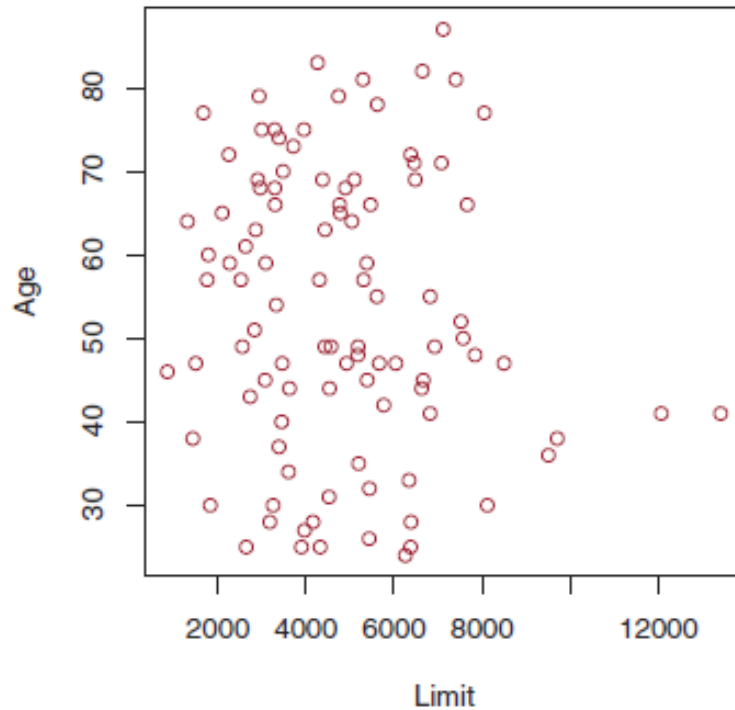
$\epsilon_1, \epsilon_2, \dots, \epsilon_n$, are uncorrelated.



Heterocedasticidad de los residuos



Multicolinealidad entre las variables explicativas



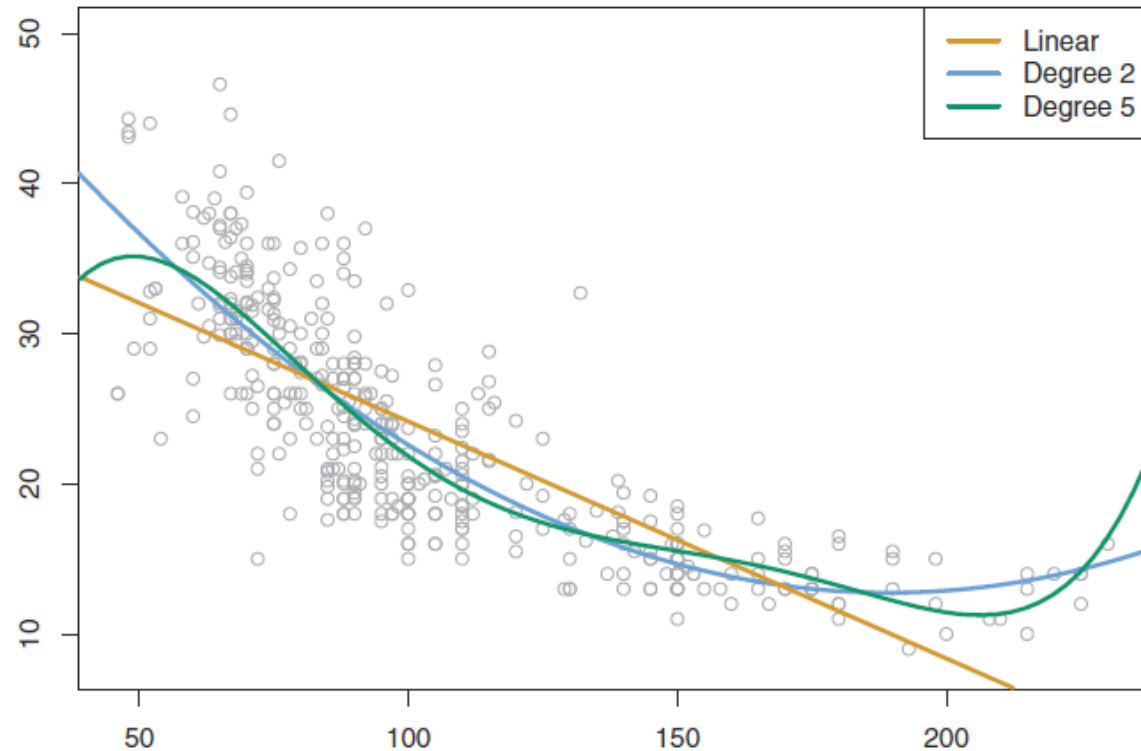
Relaciones No Lineales

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

+ *quadratic term*, X_2^2

→ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2$

polynomial regression.



Selección de Variables Predictoras

Es tan fácil añadir más variables que asumimos aportan información o tienen relación directa con la variable objetivo. Algunos de estos predictores serán significativos. ¿Quizás hay una relación, o es sólo por casualidad? El R^2 *ajustado* es una versión modificada del R^2 que ha sido ajustada para el número de predictores en el modelo. El R^2 *ajustado* aumenta sólo si el nuevo término mejora el modelo más de lo que se esperaría por casualidad y disminuye cuando un predictor mejora el modelo en menos de lo esperado por casualidad. El R^2 *ajustado* hasta puede ser negativo

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$\text{Adjusted } R^2 = 1 - \frac{RSS/(n - d - 1)}{TSS/(n - 1)}$$



Selección hacia delante

1. Llamemos M_0 al modelo nulo que no contiene predictores.
2. Para $k = 0, \dots, p - 1$:
 - a) Evaluar los $(p - k)$ modelos que se generan al incorporar un nuevo predictor a M_k
 - b) Elegir como el mejor de estos modelos $(p - k)$ modelos, que pasará a ser el M_{k+1} , a aquel que tenga el R^2 más alto.
3. Seleccionar el modelo con el mejor R^2 ajustado entre los modelos M_0, M_1, \dots, M_p

Selección hacia Atrás

1. Llamemos M_0 al modelo completo que contiene a los todos los p predictores
2. Para $k = p, p-1, \dots, 1$
 - a) Evaluar los k modelos que contengan a todos menos a un predictor de M_k , considerando $k-1$ predictores
 - b) Elegir como el mejor de estos k modelos, que pasará a ser el M_{k+1} , a aquel que tenga el R^2 más alto
3. Seleccionar el modelo con el mejor R^2 ajustado entre los modelos M_0, M_1, \dots, M_p

