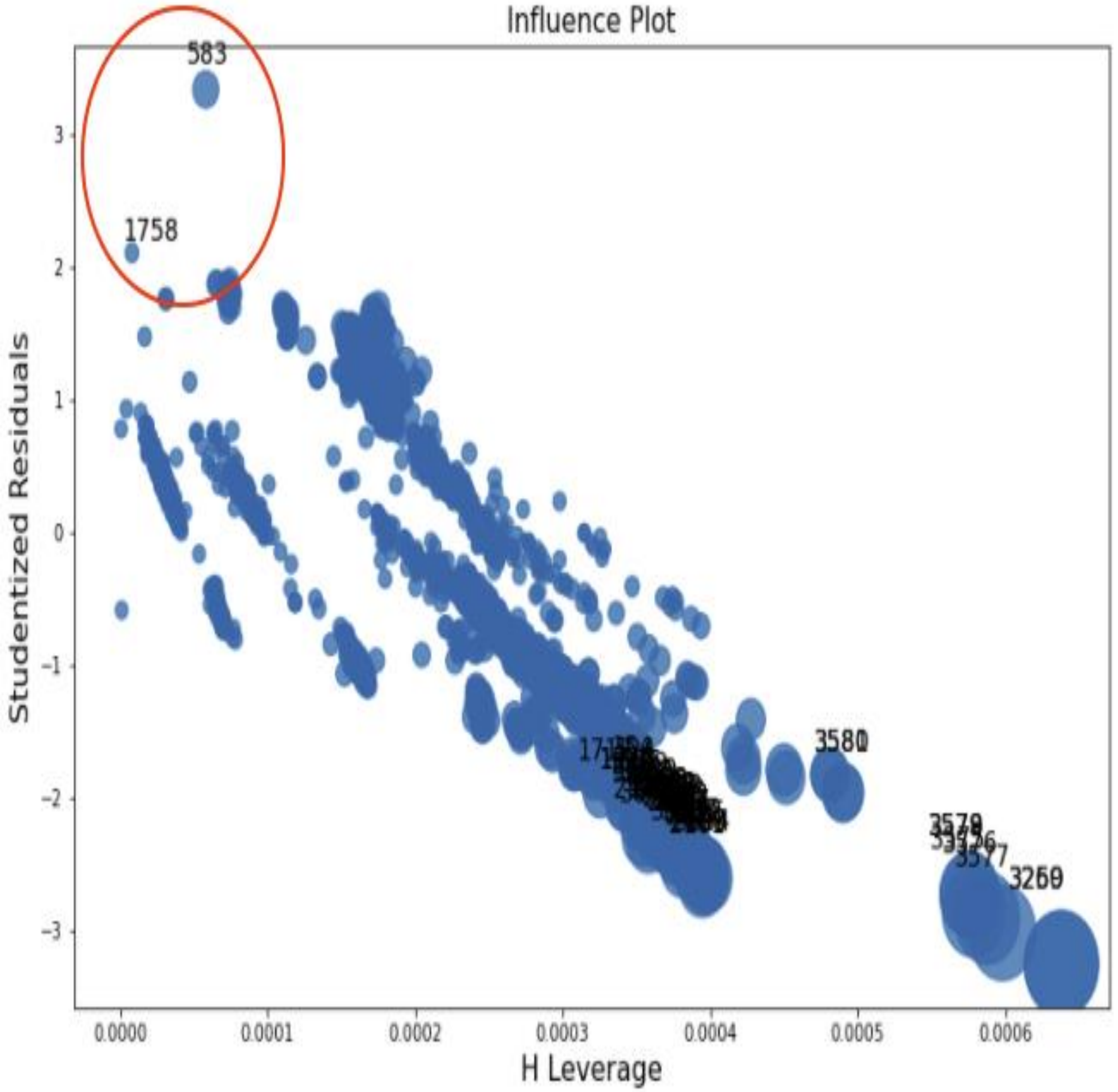


Medidas remediales en la RLM

Óscar Centeno Mora

Preámbulo

- En el uso de la RLM, seleccionamos los predictores o variables dependientes, volvimos a estimar la RLM con las variables seleccionadas, analizamos las condiciones y / o los supuestos mediante diagnósticos, y resulta que no poseemos un modelo que cumple con los requerido…

- ¿Qué es lo que debemos realizar?

- Se podría: no hacer nada, aplicar otras técnicas de estimación, ver en qué se cumple y en que no, y pues tomar decisiones, o aplicar ciertas medidas remediales.

- El aplicar una o varias “medidas remediales” podría ser una alternativa, cuando esta así lo amerite, pero se debe de tener plena conciencia en que haya claridad a la hora explicar la RLM (Ej: las transformaciones).

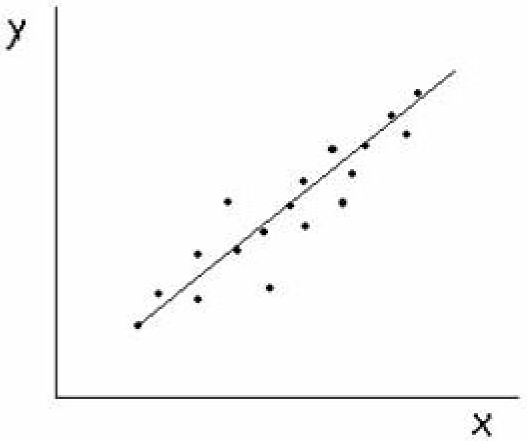
Preámbulo

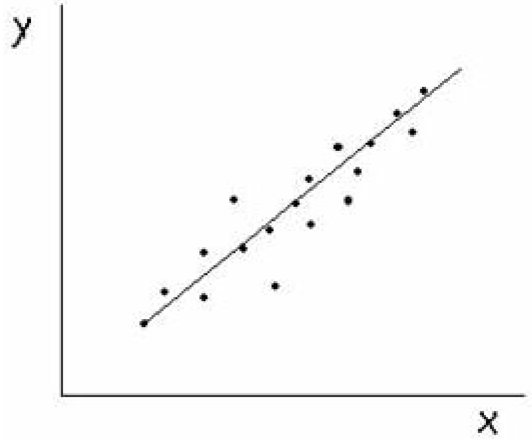
<u>Regresión bivariada</u>

Una variable dependiente (Y)
Una variable independiente (X)

$$Y \sim X$$

$$\begin{bmatrix} Y1 \\ Y2 \\ \cdots \\ Y_n \end{bmatrix} \sim \begin{bmatrix} X1 \\ X2 \\ \cdots \\ X_n \end{bmatrix}$$





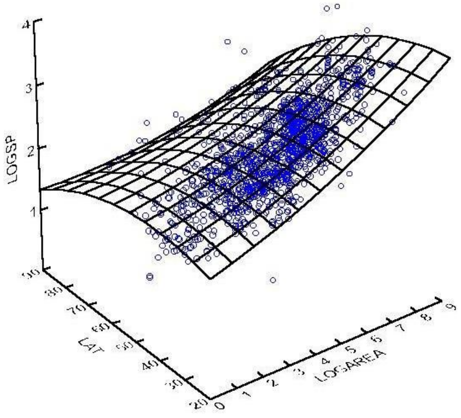
<u>Regresión multivariada</u>

Una variable dependiente
Dos o más variables independientes

$$Y \sim X$$

$$\begin{bmatrix} Y1 \\ Y2 \\ \cdots \\ Y_n \end{bmatrix} \sim \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

¿ cómo podemos obtener un mejor ajuste de la RLM?



Índice

Índice

Introducción

La estimación de la RLM no resultó ser la más adecuada, pero antes de movernos a otros enfoques de estimación, queremos ver si podemos "remediar" algunas condiciones y así seguir optando por la RLM. Y surge la duda:

¿en materia de qué se podría remediar la RLM?

Introducción

Veamos el siguiente cuadro:

Condición - Supuesto	Se prueba la…	Pero se tiene…	Se podría remediar mediante…
Normalidad	Normalidad de los residuos	No normalidad	Transformación – regresión de conteos
Variancia constante	Presencia de homocedasticidad	Heteroscedasticidad	Transformación, regresión de mínimos cuadrados ponderados.
Independencia lineal	No correlación entre las variables Xs	Multicolinealidad	Eliminar alguna variable, regresión de Ridge o de PCA.
Relación lineal	Relación lineal entre la Y y las Xs	No linealidad	Transformación.
Valores extremos - influencia	“Ausencia” o la no afectación de estos en la RLM	Afectación de casos de influencia en la RLM	Eliminar influencias, regresión a la media o robusta

Introducción

- Existen ciertas medidas remediales que podrían seguir salvando el hecho de querer permanecer con la RLM… aunque tal vez esta no sea la mejor opción.

- Según el cuadro anterior, veremos ciertas medidas como lo son, la eliminación de los valores de influencia, realizar transformacions, regresión ponderando los errores, regresión mediante el sesgamieto (Ridge), regresión a la mediana, etc.

- Es importante volver a resaltar que la regression lineal multiple, busca respetar la condición de:

$$\boxed{\varepsilon \sim N(0, \sigma^2 I)}$$

- Sin embargo, por más “curitas” que quiseramos poner, hay que sentar el presedente que tal vez, una estimación por MCO no sea lo major, que existen otros enfoques de estimación y que sin duda podrían haber otros enfoques de análisis que podrían mejorar mejorar la comprensión de la problemática puesta en causa.

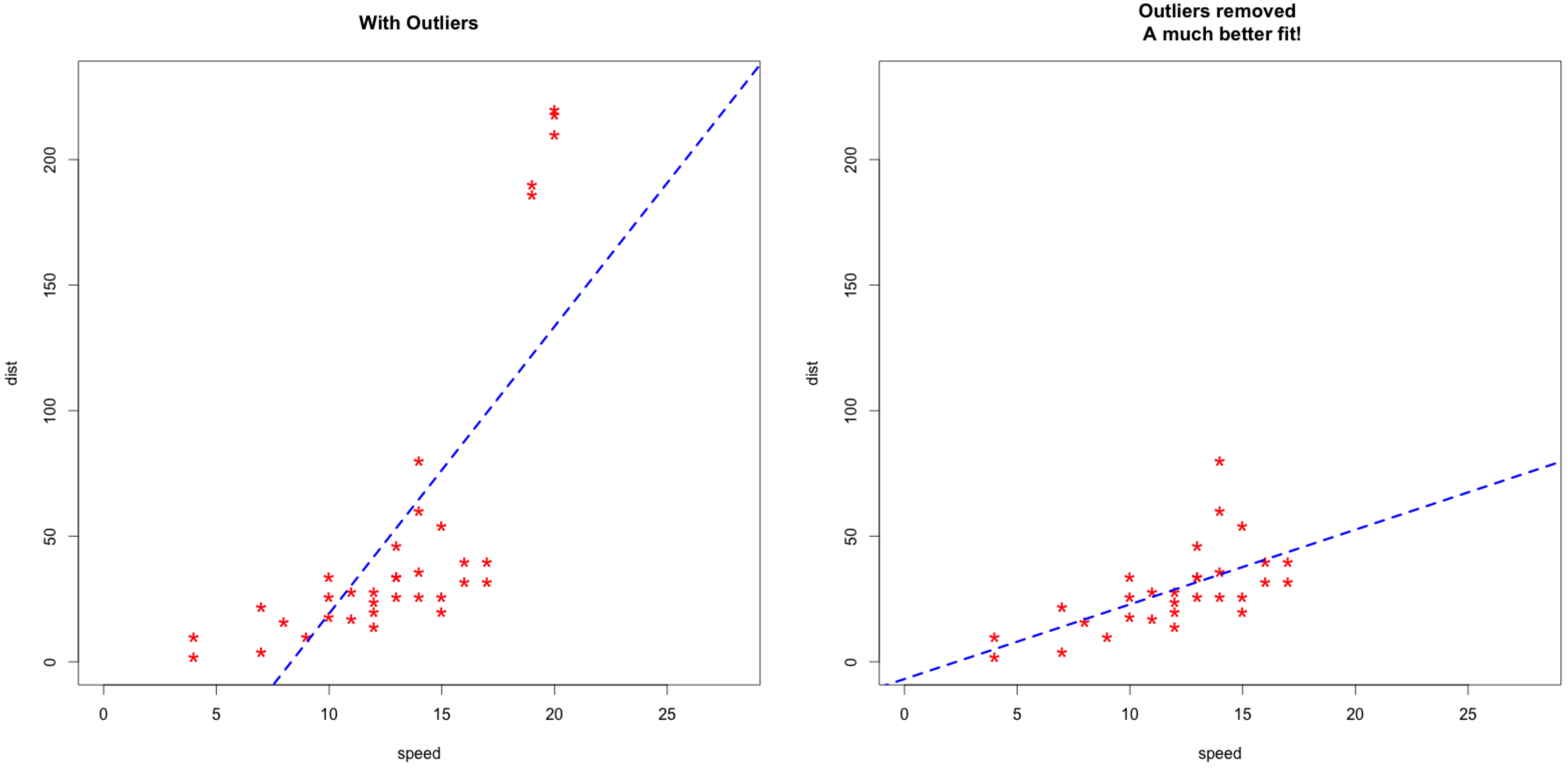
Índice

Eliminar casos – outliers

- ¿A qué nos referimos por eliminar casos de influencia o eliminar outliers en la RLM?

Eliminar casos – outliers

- Exactamente a esto:



Eliminar casos – outliers

- La eliminación de un o varios valores extremos punto de influencia puede ser a veces lo mejor: estos podrían estar distorsionando todas las condiciones de la RLM.

- Tambien, a veces los puntos de influencia suelen ser errores sistemáticos: valores por errores de las personas. Por lo que esta opción se debe de contemplar como una primera opción, en vez tratar de solucionar todos los anteriores.

- Reposa en el contexto que también NO se deben de eliminar los valores extremos o puntos de influencia, dado que estos son totalmente plausibles, y así se deben de considerar.

- En estos casos se debe de optar por incluirlos, y por ende, puede pensarse en otros métodos estimación (regresión a la media, regresión robusta, etc).

Eliminar casos – outliers

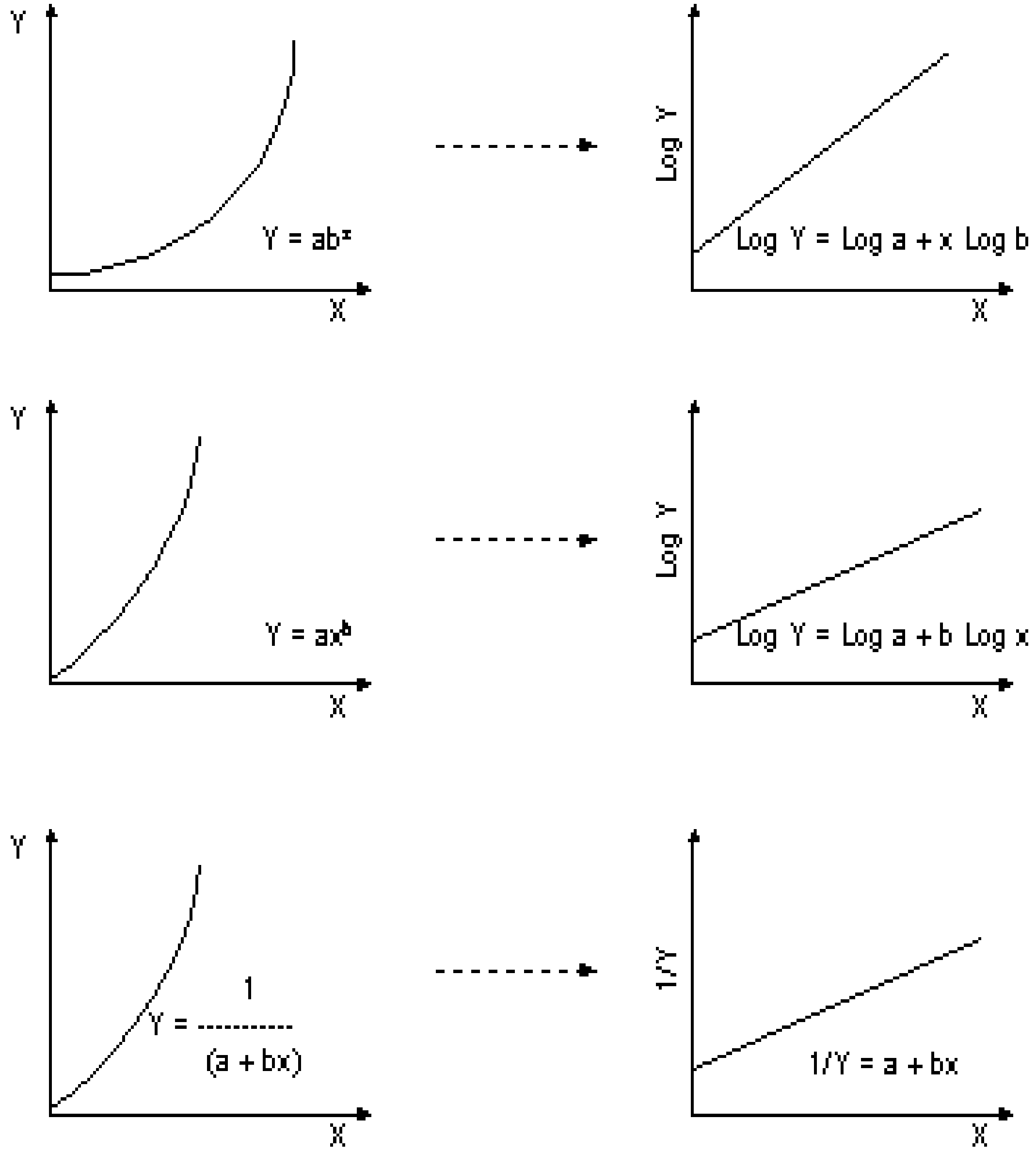
- Eliminamos los casos de influencia → volvemos a corroborar los supuestos.

- ¿Y si aún no se ha arreglado?

Índice

La transformación de variables

- ¿A qué nos referimos con transformación de la o las variables?
- A esto:
- No todas las relaciones son linealies, y pues debemos aceptar este hecho.
- Al hacer la transformación, linealizamos la respuesta en Y en las Xs.



La transformación de variables

- La transformación suele solucionar tanto los problemas de normalidad, heteroscedasticidad, y claro está, linealidad en referencia a la variable Y.

- Las transformaciones puede suceder de 3 formas: en las X, en la Y, o en ambas, dado que a veces se necesita una transformación simultánea en las X para obtener o mantener la relación lineal.

- Las transformaciones $y^{1/2}$, *log(y)* y *1/y* son apropiadas cuando la variancia crece o decrece con la respuesta, pero cada una es más severa que la anterior. La raíz cuadrada es relativamente suave y es especialmente apropiada cuando la respuesta sigue una distribución Poisson. El logaritmo es la transformación más usada. Es apropiada cuando la desviación estándar del error es un porcentaje de la respuesta. Finalmente, la transformación inversa (*1/y*) es aplicada cuando la respuesta es el tiempo hasta algún evento. Convierte tiempos por evento en tasa por unidad de tiempo. Es apropiada cuando las respuestas se apuñan cerca del cero.

- Podemos utilizar el método de Box-Cox para determinar la transformación más adecuada.

La transformación de variables

- El método de Box-Cox es una alternativa automática para encontrar una transformación para la respuesta cuando ésta no es obvia a partir de los gráficos. Se estima el parámetro λ junto con los coeficientes del modelo por el método de máxima verosimilitud. Un gráfico del perfil del logaritmo de la verosimilitud (log-likelihood) para diferentes valores de λ permite observar en qué punto alcanza su máximo. A partir de ahí se puede escoger un λ en el vecindario donde se alcanza el máximo.

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda \tau^{\lambda-1}} & \lambda \neq 0 \\ \ln(y_i)/\tau & \lambda = 0 \end{cases}$$

τ es la media geométrica

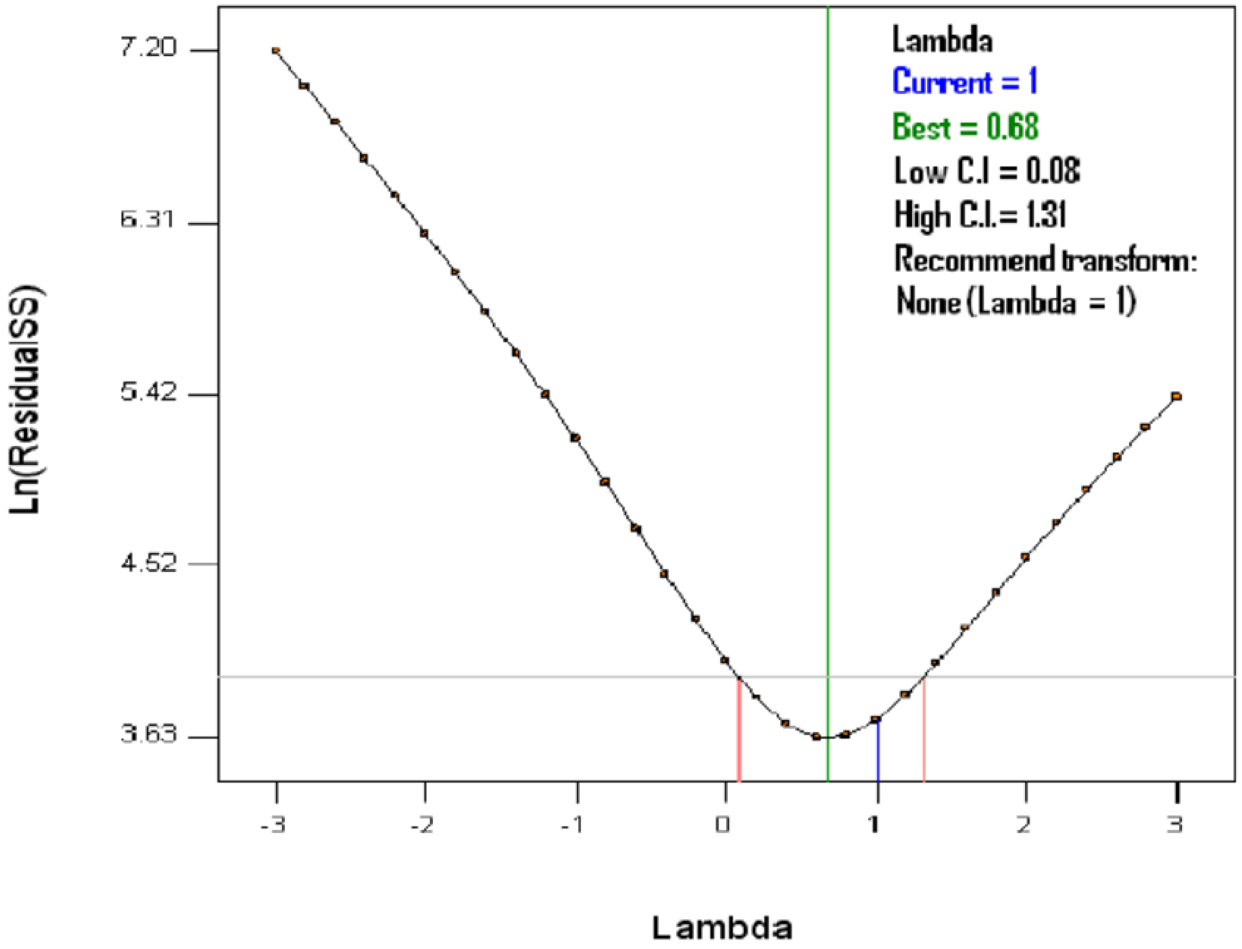
- El log-likelihood puede ser obtenido mediante: $l(\lambda) = -\frac{n}{2}\ln\left[\frac{SCE_\lambda}{n}\right] - \frac{n}{2}\left[1 + \ln(2\pi)\right]$

- La parte del log-likelihood que debe ser maximizada es: $-\frac{n}{2}\ln\left[\frac{SCE_\lambda}{n}\right]$

- El log-likelihood alcanza su máximo cuando la SCE es mínima, por lo tanto, basta estimar el modelo para diferentes valores de λ y observar para cuál valor de λ se minimiza la SCE.

La transformación de variables

- Según el gráfico de Box-Cox, se sugerirá una determinada transformación en las variables.



λ	Transformed Data
-2	y^{-2}
-1	y^{-1}
-0.5	$1/\sqrt{y}$
0	$\ln(y)$
0.5	$\sqrt{y}$
1	y
2	y^2

La transformación de variables.

- La transformación parece ser una “medida remedial” bastante atractiva, pero muchas veces es a costa de “torturar” nuestros datos.
- Otro problema se deriva de la interpretación de los coeficientes… llega a ser una labor imposible de comprender.

La transformación de variables

- Transformamos las variables → volvemos a corroborar los supuestos.
- ¿Y si aún no se ha arreglado?

Índice

La regresión de Rigde → contra la multicolinealidad

- Antes de abordar la regresión de Rigde, se puede plantear introducir a la RLM las variables que no posean una alta colinealidad entre ellas.

- Si las variables A y B son combinaciones lineales de cada una, pues se puede plantear únicamente introducir A, o solo introducir B, y de forma más práctica eludir el problema de la multicolinealidad.

- Y si no fuera posible, entonces se debe recurrir a métodos como la regresión de Rigde, la componentes principales, o la regresión por PLS (Partial Least Squares).

- Abordaremos únicamente la regresión de Ridge.

La regresión de Rigde → contra la multicolinealidad

- La idea de aplicar la regresión de Rigde es obtener estimadores sesgados con alta precisión los cuales tienen una mayor probabilidad de estar cerca del parámetro que un estimador insesgado pero impreciso. En otros términos, los coeficientes estimados serán sesgados, pero esto producirá una ganancia dado que se eliminará el problema de la multicolineadad.

- Los coeficientes se obtienen introduciendo en las ecuaciones normales de mínimos cuadrados un constante de sesgo c mayor que cero. Sean las siguientes postulaciones.

$$Y_i^{'} = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{s_Y}\right)$$

$$X_{ik}^{'} = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \bar{X}_k}{s_k}\right)$$

X — Matriz nx(p-1) con los predictores transformados X′.

$X'X = r_{XX}$ — Matriz de correlación de las X.

$X'Y = r_{XY}$ — Coeficientes de correlación de Y con cada X.

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y = r_{XX}^{-1} r_{XY}$$

(p-1) coeficientes estimados sin intercepto.

La regresión de Rigde → contra la multicolinealidad

- Introducimos una constante c, la cual será la responsable del sesgamiento en los coeficientes.

$$\hat{\beta}^R = \left(r_{XX} + cI\right)^{-1} r_{XY}$$

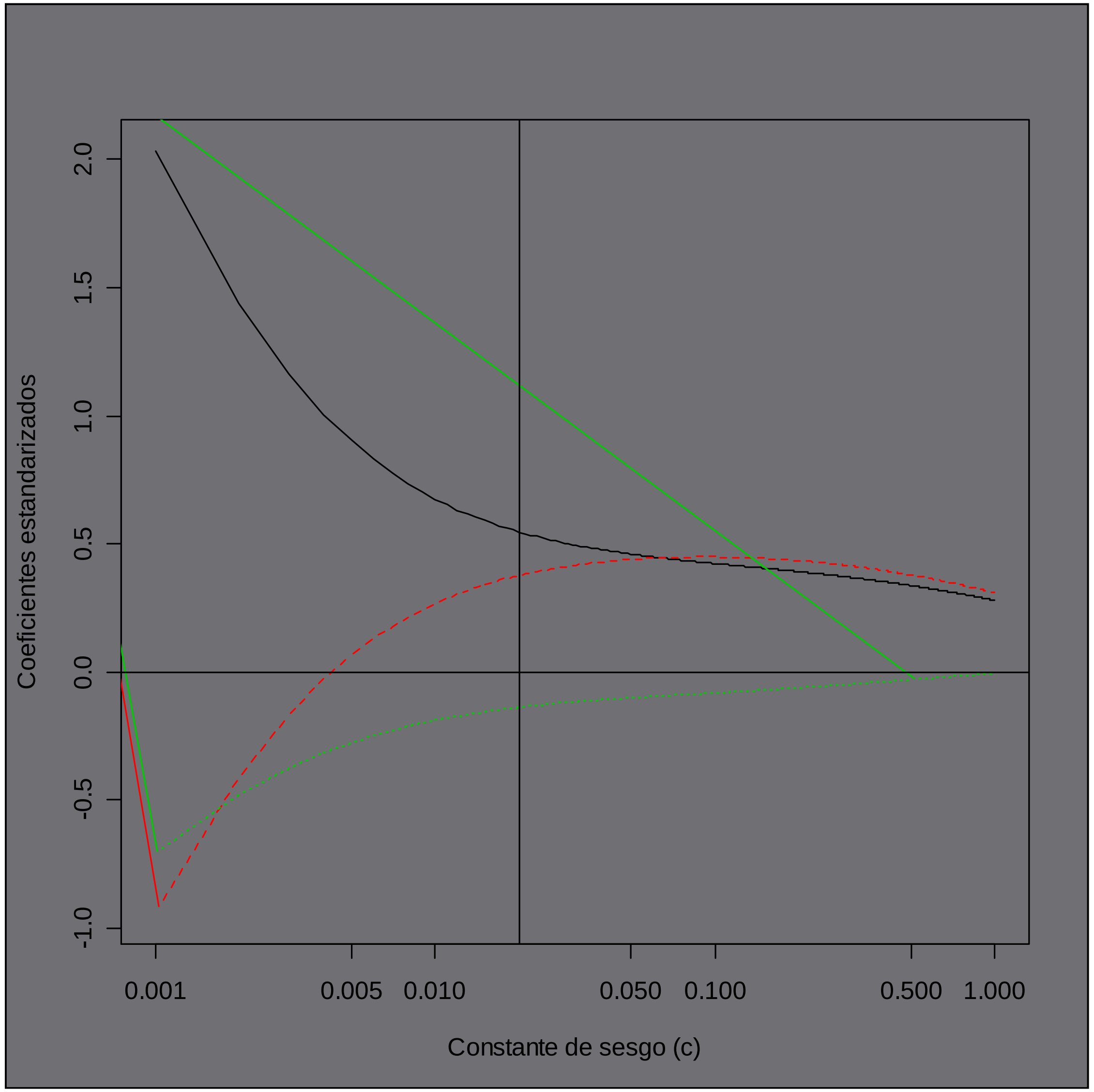
- La constante c refleja la cantidad de sesgo en los estimadores. Cuando $c = 0$ se tienen los coeficientes de mínimos cuadrados para variables estandarizadas. Cuando $c > 0$ los coeficientes son sesgados pero tienden a ser más estables (i.e., menos variables) que los de mínimos cuadrados.

- En la escogencia de la constance de sesgo c, se suele analizar un gráfico de curvas simultáneas de los diferentes coeficientes estimados para los predictores contra los valores de c (usualmente en escala logarítmica).

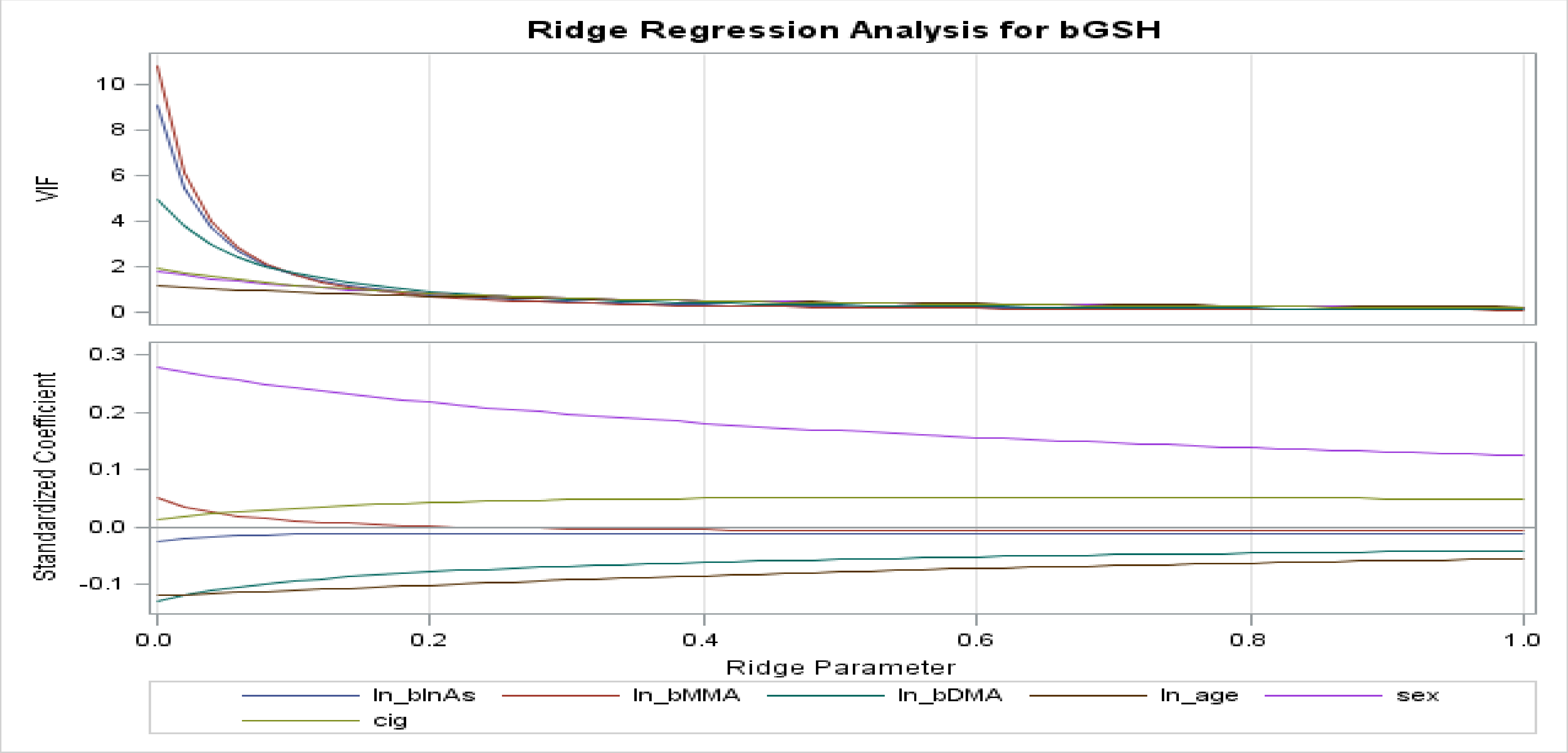
- Se escoge el menor valor de c donde se observa una estabilidad en los coeficientes estimados de todas los predictores y los VIF (factores de inflación de la varianza) son suficientemente pequeños.

La regresión de Rigde → contra la multicolinealidad

- Al usar una constante c=0.02 se logra un modelo con un R^2=0.78 mientras que el modelo original tenía un R^2=0.80. El VIF se reduce de 708, 564 y 105 a 1.1, 1.1 y 1.0 para los tres predictores, respectivamente.



La regresión de Rigde → contra la multicolinealidad



La regresión de Ridge

- Realizamos la regresión de Ridge → volvemos a corroborar los supuestos.

- ¿Y si aún no se ha arreglado?

Índice

Heteroscedasticidad → Mínimos cuadrados ponderados

- Al poseer variancia no constante, o heteroscedasticidad, en gran parte se puede deber a casos que poseen residuos desiguales.

- Una forma de remediar dicho problema de desigualdad residual es ponderando estos, y tratar de "equiparar las diferencias residuales". Esto podría solucionar el problema de la variancia no constante.

- Una forma es aplicar una regresión ponderada. Para esto debemos el peso o los ponderados de ante mano.

- Para estimar la función de varianza o de ponderación, ajustamos el modelo de regresión usando mínimos cuadrados ordinarios (sin ponderar) y se hace la regresión de los errores contra las variables apropiadas. Si hay valores extremos en los datos es mejor estimar la función de desviación estándar en lugar de la de varianza porque se ve menos afectada por los valores extremos. Los valores estimados en esta regresión sirven como estimaciones de la varianza o la desviación estándar. Las ponderaciones serán los inversos de la varianza estimada.

Heteroscedasticidad → Mínimos cuadrados ponderados

- Dependiendo de la forma de la heteroscedasticidad, se puede optar por diferentes formas de calcular u obtener los ponderados:

  - Si el gráfico de residuales contra X tiene forma de megáfono: ajustar los residuales absolutos contra X
  - Si el gráfico de los residuales contra $\hat{Y}$ tiene forma de megáfono: ajustar los residuales absolutos contra $\hat{Y}$.
  - Si el gráfico de los residuales cuadráticos contra X presenta una tendencia hacia arriba: ajustar los residuales cuadráticos contra X.
  - Si el gráfico de los residuales contra X presenta un rápido crecimiento de la varianza conforme X aumenta hasta cierto punto y luego crece más despacio: ajustar los residuales absolutos contra X y X^2.

- Aunque no vamos a cubrir todas las posibles formas de obtener los ponderadores, es importante saber que existen las anteriores recomendaciones.

- Veamos ahora el proceso de estimación de una regression por mínimos cuadrados ponderados.

Heteroscedasticidad → Mínimos cuadrados ponderados

- Paso 1: Ajustar el modelo de regresión sin ponderar y analizar los residuales.
- Paso 2: Estimar la función de varianza o desviación estándar.
- Paso 3: Usar los valores ajustados para obtener las ponderaciones:

$$w_i = \frac{1}{\left(\hat{s}_i\right)^2} \qquad W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & & \dots \\ 0 & 0 & \dots & w_n \end{bmatrix}$$

- Estimar los coeficientes de regresión usando estos pesos:

$$\hat{\beta}_w = (X'WX)^{-1} X'WY$$

Heteroscedasticidad → Mínimos cuadrados ponderados

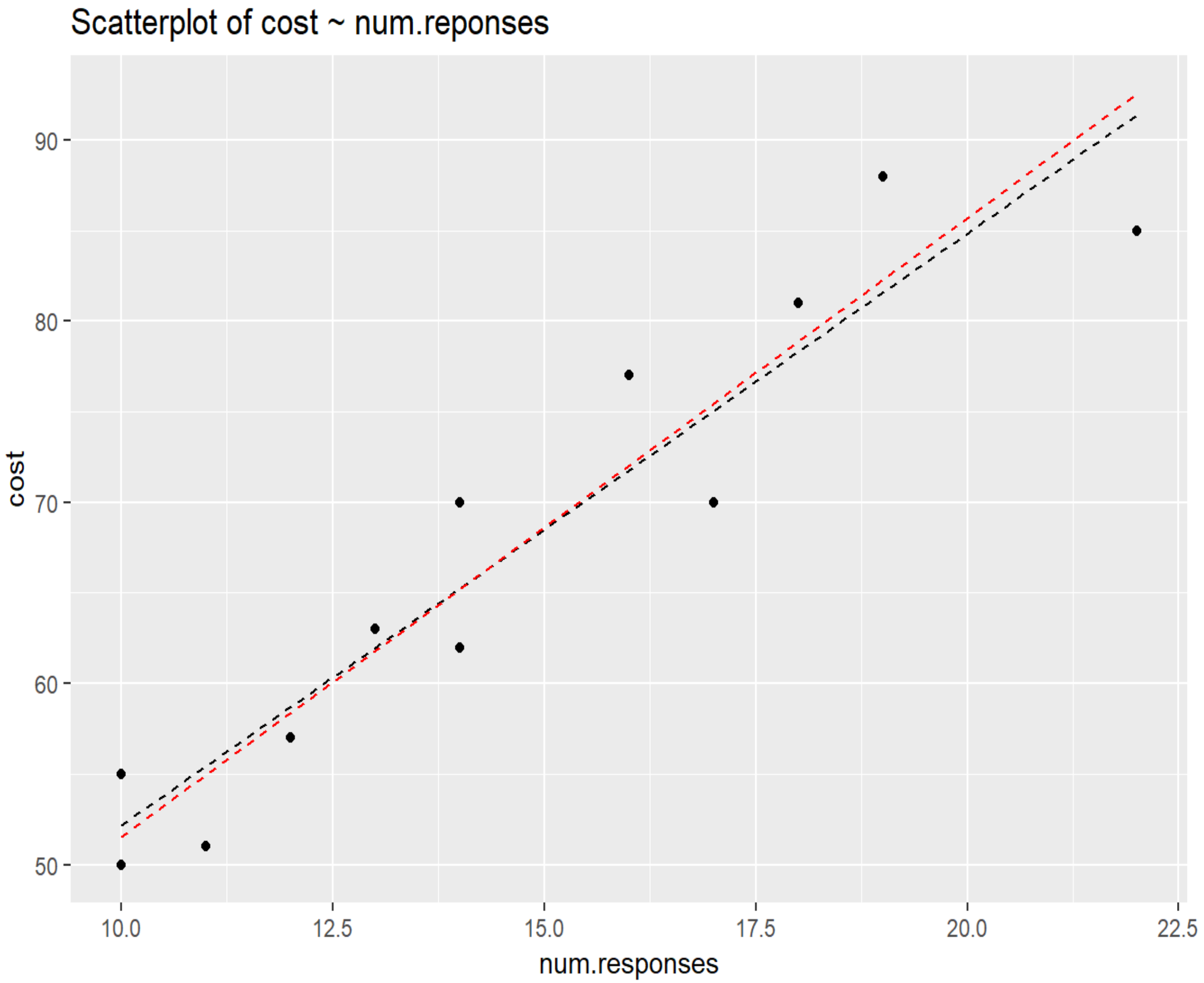
- Si las estimaciones de los coeficientes cuando se usaron ponderaciones difieren sustancialmente de los primeros coeficientes cuando no se ponderó, se sugiere iterar el proceso de ponderación usando los residuales de los mínimos cuadrados ponderados para volver a estimar la función de varianza o desviación estándar y luego obtener nuevos pesos.

- A menudo basta una o dos iteraciones para estabilizar los coeficientes de regresión.

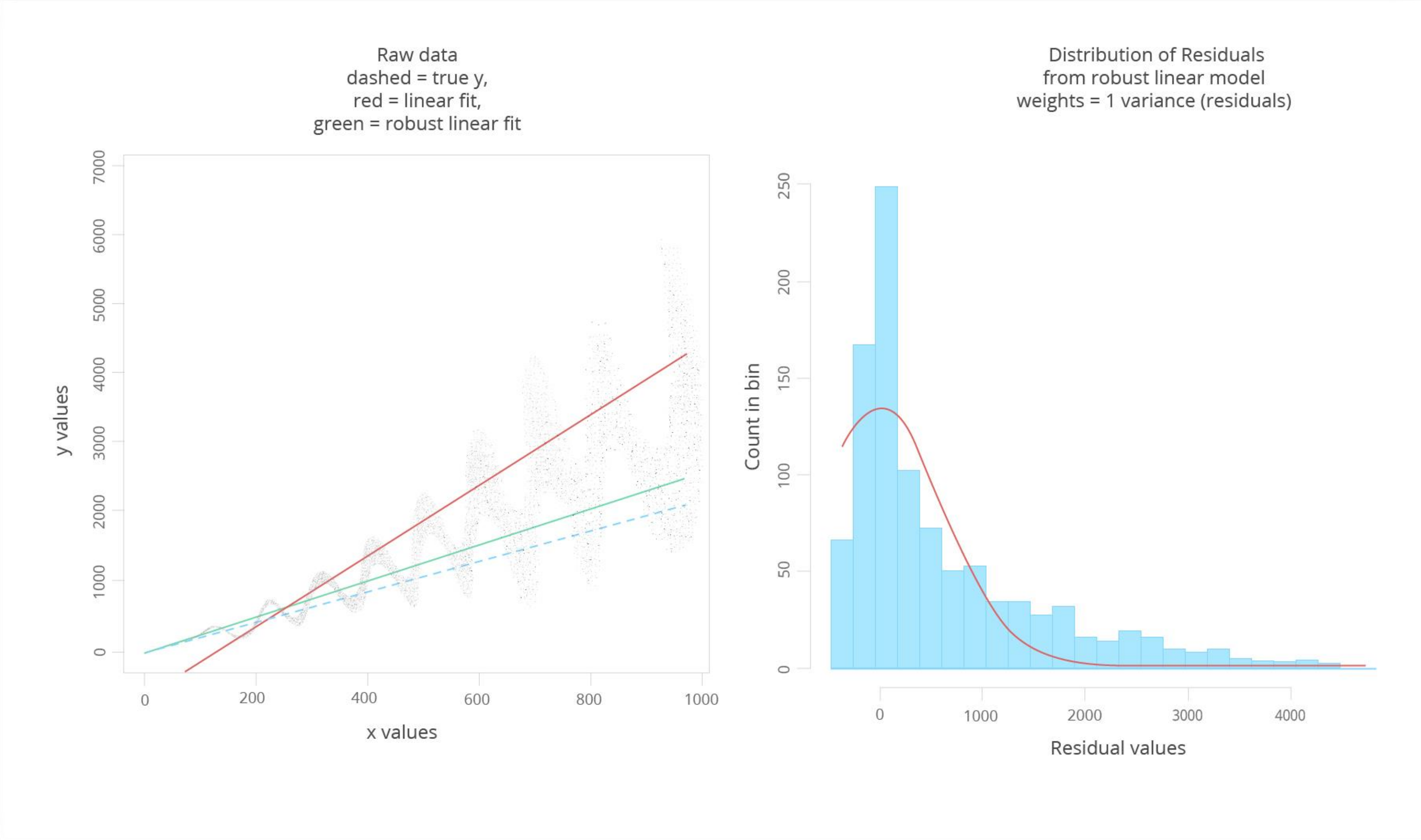
- Cuando hay poca diferencia en las varianzas de los errores el método de mínimos cuadrados ponderados no es muy útil….

- Veamos un caso gráfico

Heteroscedasticidad → Mínimos cuadrados ponderados



Heteroscedasticidad → Mínimos cuadrados ponderados



La regresión ponderada

- Realizamos la regresión ponderada → volvemos a corroborar los supuestos.

- ¿Y si aún no se ha arreglado?

Regresión robusta → valores de influencia

- En la primera parte analizamos el eliminar puntos de influencia o valores extremos de influencia. Sin embargo, sucede que un análisis más detallado de la situación no nos permite eliminar el valor extremo, dado que este es justificable: es un evento que puede suceder.

- Eliminar valores extremos puede provocar a su vez la perdida de precisión y la perdida de información… por lo que la opción más conveniente, no suele ser a veces la mejor.

- Para proceder con algún tipo de regresión, se podría aplicar una regresión a la mediana (regresión cuantílica), o se podría aplicar una regresión robusta, la cual veremos a continuación.

- No se debe menospreciar la regresión cuantílica, dado que para datos económicos, con presencia de asimetría positiva, me parece que es una excelente elección.

Regresión robusta → valores de influencia

- La regresión robusta busca reducir el peso de los valores de influencia, en comparado con los estimadores de mínimos cuadrados ordinarios, lo cuales no realiza ningún estrategia en este sentido…

- Una de las ventajas, además de aminorar el peso de las influencias, es aportar un mejor ajuste para la mayoría de los datos sin necesidad de desechar información.

- Para aproximaciones de regresiones automáticos, es una gran elección: por ejemplo, un instrumento de medición complejo usado para exámenes médicos debe calibrarse cada vez que se use y no hay tiempo para identificar valores extremos.

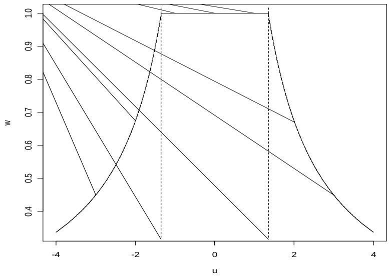
- Para nuestro caso, estudiaremos la IRLS: Iteratively reweighted least squares. Esta aproximación utiliza el procedimiento de mínimos cuadrados ponderados para reducir la influencia de valores extremos. En lugar de usar los pesos basados en las varianzas de los errores, usa pesos basados en la lejanía de cada caso, medida por el residual de cada caso. Se revisan los pesos en cada iteración hasta que se obtenga un ajuste robusto.

Regresión robusta → valores de influencia

- En el uso de la regresión robusta, se suelen utilizar dos funciones ponderadoras:

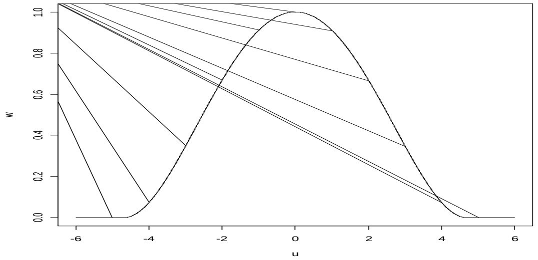
<u>Huber</u>:

$$w_i = \begin{cases} 1 & |u_i| \le 1.345 \\ \dfrac{1.345}{|u_i|} & |u_i| > 1.345 \end{cases}$$



<u>Bisquare</u>:

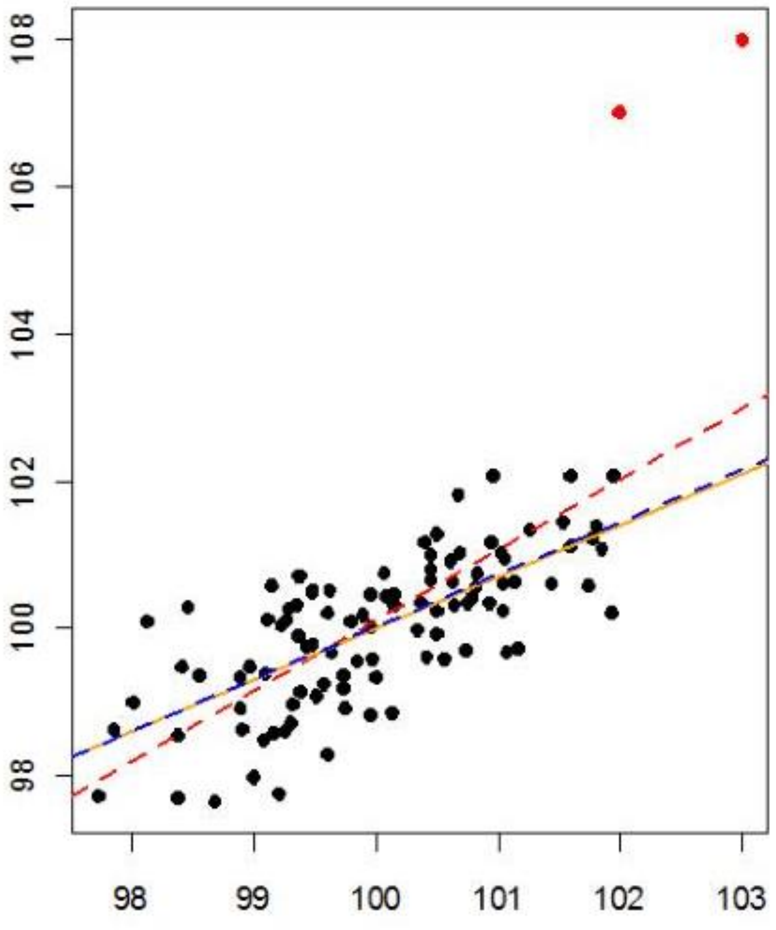
$$w_i = \begin{cases} \left[1-\left(\dfrac{u_i}{4.685}\right)^2\right]^2 & |u_i| \le 4.685 \\ 0 & |u_i| > 4.685 \end{cases}$$



- El valor u_i denota el residual escalado. Los valores 1.345 y 4.685 se llaman constantes de "tunning" que se escogen para hacer el procedimiento IRLS 95% eficiente para datos generados por el modelo de regresión con errores normales. El peso w_i baja conforme el residual escalado absoluto crece.

Regresión robusta → valores de influencia

- Veamos una aplicación de la regresión robusta:



Data set used

Number of data n=102

100 data

~bivariate normal (cor=0.7, var=1)

2 data

artificial outliers

Results of regression Analysis

OLS: α 4.74 β 0.95

OLS without outliers:

α 30.0 β 0.7

IRLS: α 29.10 β 0.71

La regresión robusta

- Realizamos la regresión robusta → volvemos a corroborar los supuestos.

- ¿Y si aún no se ha arreglado?


EXTRA
EXTRA
EXTRA

Regresión a la mediana: regresión cuantílica

¿Por qué, en problemas con asimetrías o valores extremos, vamos a preferir la mediana en vez del promedio ?

Regresión a la mediana: la regresión cuantílica

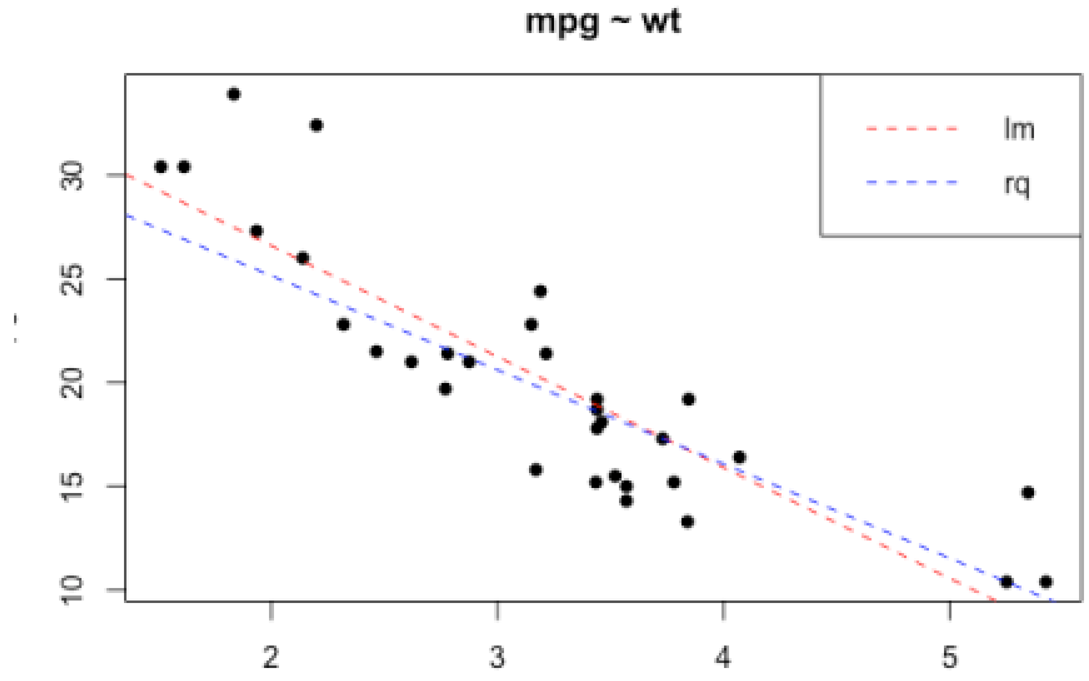
- De forma univariada, el promedio no suele ser la medida recomendada en la descripción de la información ante una distribución asimétrica.

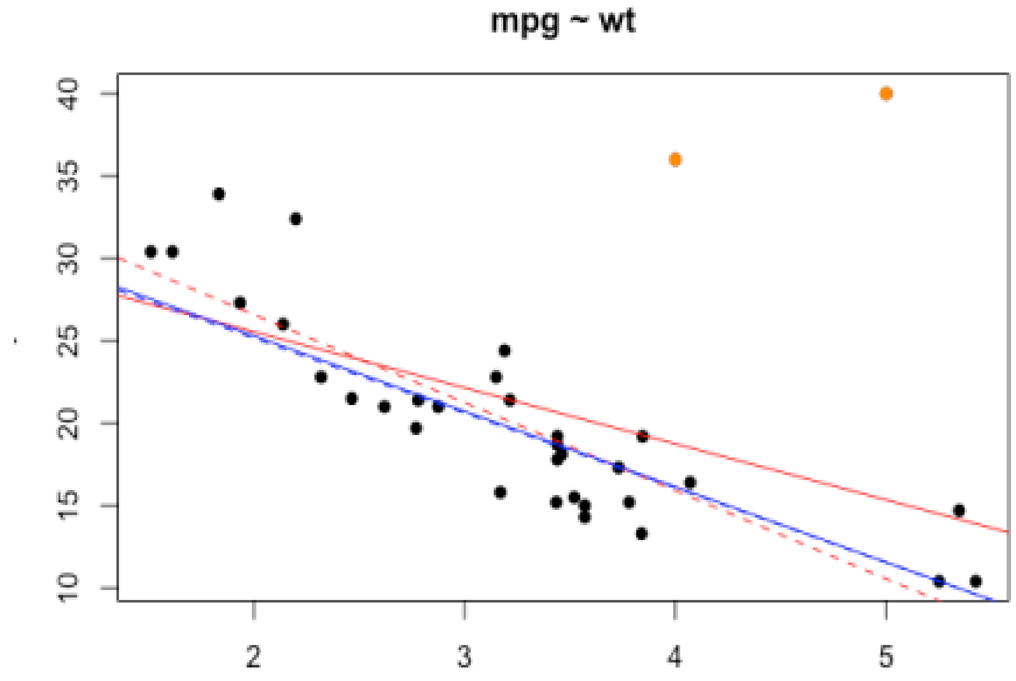
- Recordemos que los valores extremos son los causantes de poseer distribuciones asimétricas, y en tales situaciones se prefiere se prefiere la mediana como medida de posición central.

- Lo mismo sucede en un contexto de regresión: al haber uno o diversos valores extremos, el promedio no suele ser la mejor medida a referenciar, y se puede pensar en utilizar la mediana en la estimación de los coeficientes de regresión.

- La regresión cuantílica permite estudiar el impacto de las variables independientes en diferentes cuantiles de distribución de variables dependientes y, por lo tanto, proporciona una imagen completa de la relación entre Y y X, además esta es robusta a valores atípicos en observaciones de Y, y la La estimación y las inferencias están libres de distribución.

Regresión a la mediana: la regresión cuantílica





Índice

Otros enfoques

¿Nos debemos quedar únicamente con el método de OLS o MCO ?

Otros enfoques

- La respuesta es un inminente NO.

- Es imposible tartar de complacer siempre el supuesto que le hicimos a los residuos:

$$\boxed{\varepsilon \sim N(0, \sigma^2 I)}$$

- Además, en la vida existen relaciones más allá de la linealidad. Hay multiples funciones de probabilidad, y por lo tanto los residuos podrían optar por distrubuirse de otra forma más allá de la normalidad.

- También, existen las relaciones NO lineales, en donde los métodos más avanzados tratan este tipo de problématica.

- Se deben ahora seguir estudiante modelos generalizados, para seguir conociendo otros métodos de regresión.

Conclusión

- El presente capítulo estudió las medidas remediales en la RLM.

- Se analizaron formas para abarcar el caso con falores de influencia, falta de normalidad, la heteroscedasticidad, la multicolinealidad, así como el tener que estimar regresiones con puntos de influencia.

- El mundo de los OLS o MCO es muy restrictivo, y se deberá estudiar los modelos lineales generalizados, y tener en consideración la posibilidad de relaciones no lineales también.


The End