SEGUNDO CONTROL（ESTUDIO SOBRE CONSUMO DE AGUA）

Los_mas_aplicados

158228

158406

155899

1. Utilice el *Método Delta* (aproximación por expansión por series de Taylor, propagación de error), para justificar que si

$$\sigma_y^2 \propto [E(y)]^3$$

$Y = \frac{1}{\sqrt{y}}$, es la transformación que estabiliza la varianza. Esto es, $\sigma_Y^2 \propto constante$.

Demostración

El *Método Delta* nos garantiza que dada una variable aleatoria con esperanza y varianza finita y una función $h(x)$ al menos doblemente diferenciable tenemos que:

$$E[h(y)] \approx h(\mu_y) + \tfrac{h''(\mu_y)}{2}\sigma_y^2$$
$$Var(h(y)) \approx (h'(\mu_y))^2\sigma_y^2$$

Sea $h(x) = \frac{1}{\sqrt{x}} = x^{\frac{-1}{2}}$ y y una $v.a$ con esperanza y varianza finitas. Como $h(x)$ es doblemente diferenciable en los reales positivos, tenemos por el *Método Delta* que:

$$Var(\tfrac{1}{\sqrt{y}}) \approx (-\tfrac{1}{2\sqrt{\mu_y^3}})^2\sigma_y^2 \propto (-\tfrac{1}{2\sqrt{\mu_y^3}})^2\mu_y^3 \equiv 1$$

Esto es, $\sigma_Y^2 \propto k$

2. Muestre que la transformación de potencia Box-Cox

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y}\log y & \lambda = 0 \end{cases}$$

como función de λ, es continua en $\lambda = 0$. En la expresión anterior, $\dot{y} = (\prod_{i=1}^n y_i)^{\frac{1}{n}}$ es el promedio geométrico de las respuestas y_i.

Si tomamos el límite cuando $\lambda \to 0$ nos damos cuenta que tenemos una indeterminación del tipo "$\frac{0}{0}$". Al considerar a $f(\lambda) = y^\lambda - 1$ y $h(\lambda) = \lambda\dot{y}^{\lambda-1}$ es fácil observar que ambas son de al menos clase C^1. Con esto en mente, podemos aplicar la regla de L'Hopital en cuyo caso obtenemos

$$\lim_{\lambda\to 0}\frac{y^\lambda - 1}{\lambda\dot{y}^{\lambda-1}} = \lim_{\lambda\to 0}\frac{y^\lambda \log y}{\dot{y}^{\lambda-1} + \lambda\dot{y}^{\lambda-1}\log\dot{y}} = \frac{\log y}{\dot{y}^{-1}} = \dot{y}\log y$$

Por lo tanto, $y^{(\lambda)}$ es continua en $\lambda = 0$.

3. Reporte sobre el consumo de agua

   En el presente reporte se analizará el consumo de agua de cierta ciudad con respecto a su consumo de energía. Se iniciará el análisis con una visualización del conjunto de datos. Posteriormente se ajustará un modelo lineal a partir de los datos y por último se validará el modelo. Como se mostrará a continuación, será necesario considerar distintos modelos lineales transformando la variable de salida (el consumo de agua). Dicha transformación sera determinada a partir de los resultados del análisis de residuales.

   Gráfica de los datos

   La fig.1 es una gráfica de los datos en donde cada observación se representa como un punto en el plano, el eje horizontal corresponde al consumo de energía y el vertical al consumo de agua. Se puede observar una correlación positiva entre las variables. Además, parece ser que, conforme se avanza en la dirección positiva del eje horizontal, los datos se muestran más dispersos verticalmente. Lo anterior puede ser un indicio de que no se cumple el supuesto de homocedasticidad.

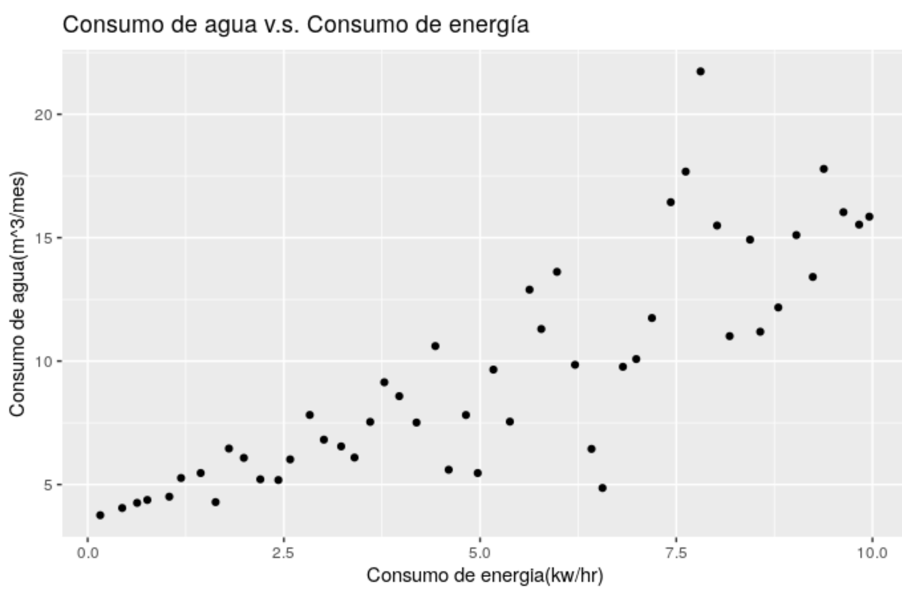

   Figure 1: Consumo de agua v.s. Consumo de energía

   Primer ajuste del modelo lineal

Se procederá ajustando un modelo lineal. En la fig.2 se muestran los datos y la recta ajustada por medio de mínimos cuadrados.

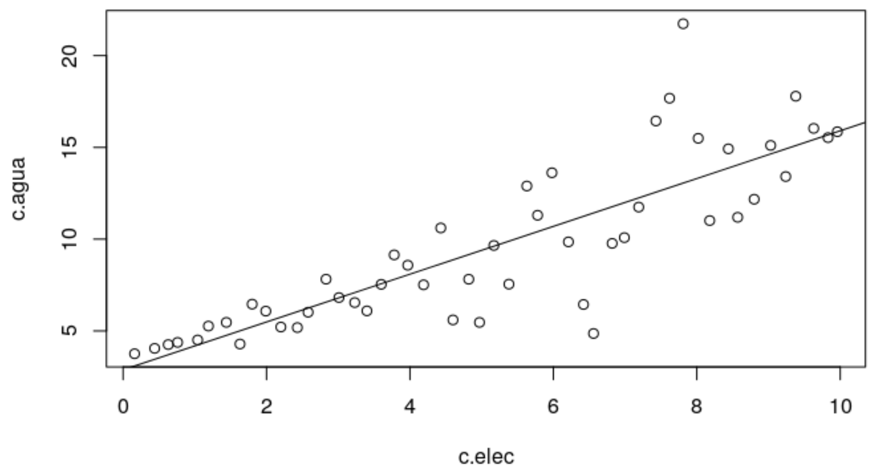

Figure 2: Primer modelo lineal ajustado

Análisis de residuales

En la fig.3 se despliega la gráfica de residuales contra los valores ajustados. Aproximadamente en el intervalo [8, 16] del eje horizontal, los puntos se muestran más dispersos verticalmente. Empieza la dispersión a crecer hasta alcanzar un aparente máximo alrededor del 13 y posteriormente parece disminuir de nuevo. En el caso de homocedasticidad los puntos no deberían mostrar ningún patron discernible. En este caso parece ser que el supuesto no se cumple.

Transformación Box-Cox

Se intentará enmendar la aparente falta de homocedasticidad de los residuales por medio de la transformación de Box - Cox. Así, construiremos la regresión lineal para la respuesta transformada por una λ en el intervalo [-3,3]. En la fig.4 se muestra la región donde se minimiza la suma de residuales al cuadrado.

Además , se muestra la recta al nivel en donde se intersecta con el intervalo de confianza al nivel 90 por ciento.

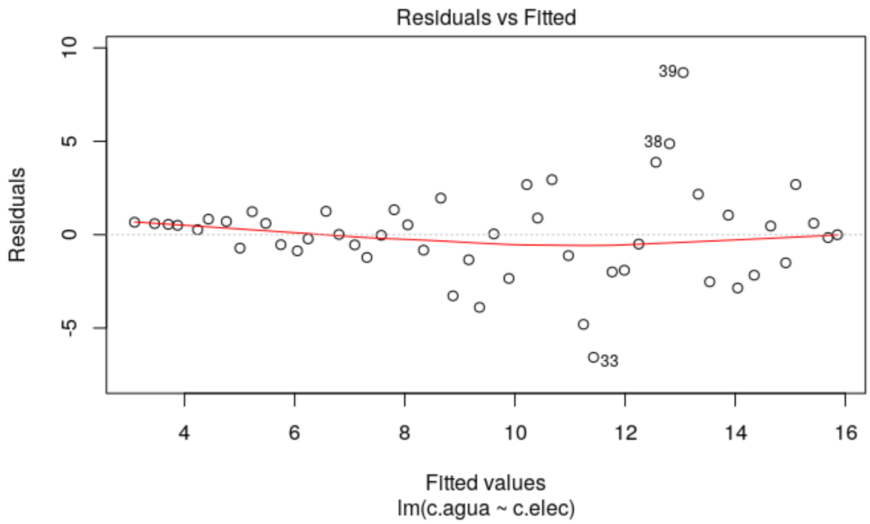

Figure 3: Respuesta ajustada v.s. Residuales del primer modelo

$$\text{SC*} = \text{SCres}(\lambda^*)(1 + t^2/\text{n-2})$$

Con t^2 el cuantil correspondiente a 0.95 de probabilidad de una variable t de student con n-2 grados de libertad y λ^* el valor que minimiza la suma de residuales. En este caso, el intervalo de confianza calculado fue [-0.455, 0.16]. Debido a que el 0 se encontraba dentro del intervalo de confianza, se decidió utilizarlo como valor de λ, pues correspondía al modelo más simple.

Una vez que se determinó el valor de λ para la transformación , se prosiguió ajustando una nueva recta que se muestra en la fig.5.

Por último se realizó el análisis de residuales para el modelo con transformación. En la fig.6 se puede observar la nueva distribución de los residuales. Ya no existe un patrón tan marcado como en la primera regresión, a pesar de lo cual se llega a notar una menor dispersión en los extremos horizontales de la gráfica.

Intervalos de confianza para la respuesta media

Para construir los intervalos de confianza para el consumo medio y la predicción del consumo fue necesario utilizar el primer modelo sin transformación de la respuesta. Los intervalos de confianza que se construyeron

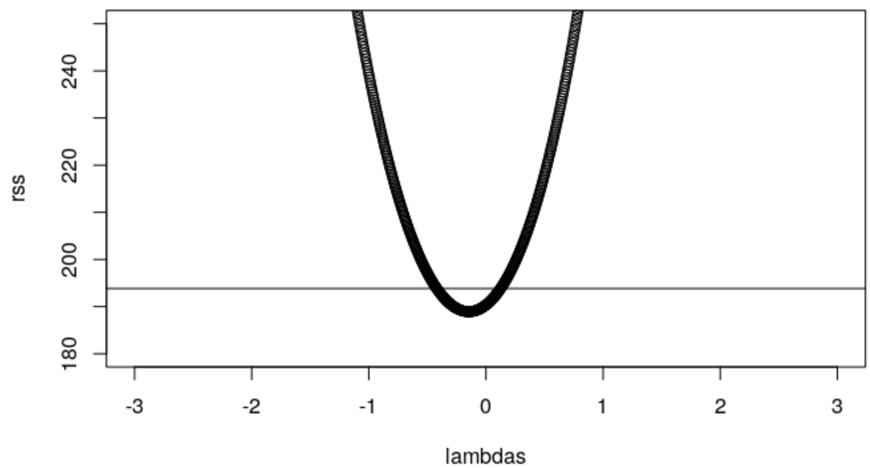

Figure 4: Rss(λ) vs λ

en clase corresponden a la respuesta utilizada en la regresión , por lo que invertir la transformación de Box-Cox hubiera distorsionado los intervalos.

Los intervalos utilizados se componen de la estimación puntual para la respuesta media y la predicción sumados y restados por el error estandar ponderado por el cuantil de orden 0.9 y 0.95 (correspondientemente) de una variable t de student con n-2 grados de libertad.

Así, un intervalo al nivel de 90 por ciento de confianza para el consumo medio correspondiente a un consume de energía de 7.57 kw/hr es [11.97,13.51]. Un intervalo de predicción al nivel de 95 por ciento de confianza para una generación de energia de 5.1 kw/hr es [4.53, 14.52]

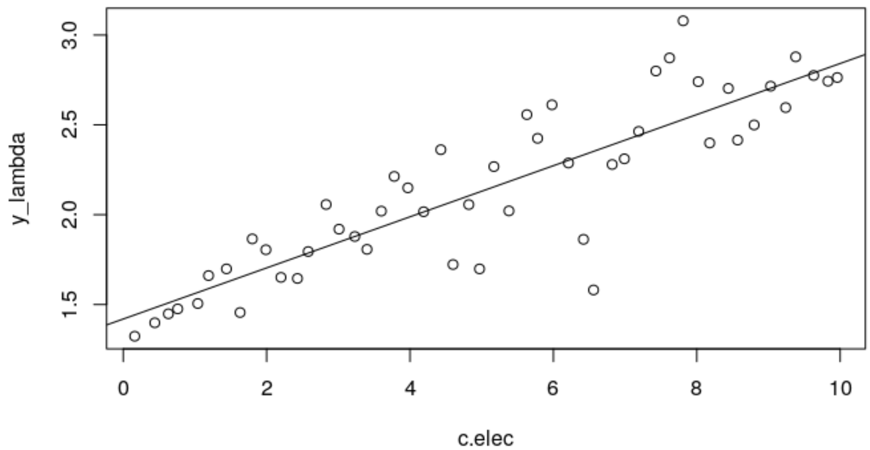

Figure 5: y^{λ} vs C.Elec y linea de regresión

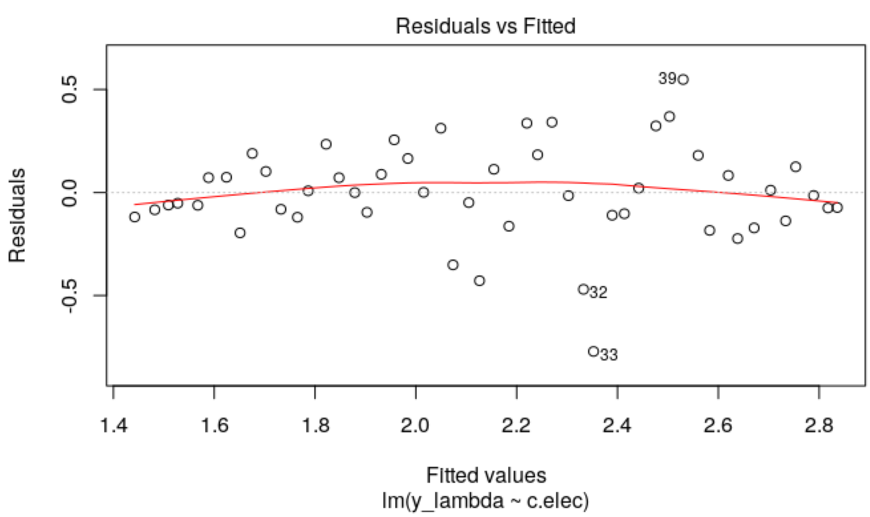

Figure 6: Residuales v.s valor ajustado con transformación