CAPITULO 6

TRANSFORMACIONES

Existen problemas para los que se conoce que la función media $\mathrm{E}[Y/X]$ es lineal.. Algunas veces, la función media se determina a partir de la teoría, tal como la data Fisico de la Sección 5.1.1. Otras veces no existe teoría disponible para decidir sobre la forma correcta de la función media y cualquier forma paramétrica que se use es poco más que un *aproximación* ya que se espera que sea adecuada al problema. Reemplazar la variable respuesta, los predictores o ambos por transformaciones no lineales de los mismos constituyen una herramienta importante que permite al analista extender el número de problemas para los que la metodología de regresión lineal es apropiada. Lo anterior permite formular dos preguntas: ¿Cómo escoger estas transformaciones? ¿Cómo decidir si un modelo aproximado es adecuado para la data?

7.1 TRANSFORMACIONES Y DIAGRAMAS DE DISPERSION

El propósito más frecuente de las transformaciones es lograr que una función media sea lineal en la escala transformada. En problemas con solo un predictor y una variable respuesta, la función media puede visualizarse en un diagrama de dispersión de tal forma la transformación seleccionada de cómo resultado un diagrama de dispersión con una función media aproximadamente lineal. Con muchos predictores, la selección de las transformaciones puede ser más difícil, así como el criterio usado por lo que se considera, a continuación, el primer caso. Se busca una transformación de tal forma que si X es el predictor *transformado* y Y es la respuesta *transformada*, entonces la función media en la escala transformada es

$$\mathrm{E}[Y/X = x] \approx \beta_0 + \beta_1 x$$

donde se usa "$\approx$" en lugar de "$=$" para indicar que la relación podría ser una aproximación no necesariamente verdadera.

La Figura 7.1 muestra el gráfico del peso del cuerpo, PesoCuerpo, en kilogramos y el peso del cerebro, PesoCerebro, en gramos para 62 especies de mamíferos (Allison y Cicchetti, 1976) que se encuentra en la data Mamiferos. Además de los tres puntos separados para dos especies de elefantes y los humanos, la distribución desigual de los puntos esconde cualquier información útil acerca de la media de PesoCerebro, dado PesoCuerpo. En cualquiera de los casos, existe poca evidencia para usar una función media lineal. El rango de ambas variables va desde especies pequeñas cuyos cuerpos pesan pocos gramos hasta animales enormes por encima de los 6600 kilogramos. Las transformaciones pueden ayudarnos en este problema.

```
Mamiferos <- read.table(file = "http://tarwi.lamolina.edu.pe/~clopez/Regresion/Mamiferos.txt", header = T)
attach(Mamiferos)
plot(PesoCuerpo,PesoCerebro,xlab="Peso del cuerpo (kg)",ylab="Peso del cerebro (g)")
identify(PesoCuerpo,PesoCerebro,labels=Especie)
```

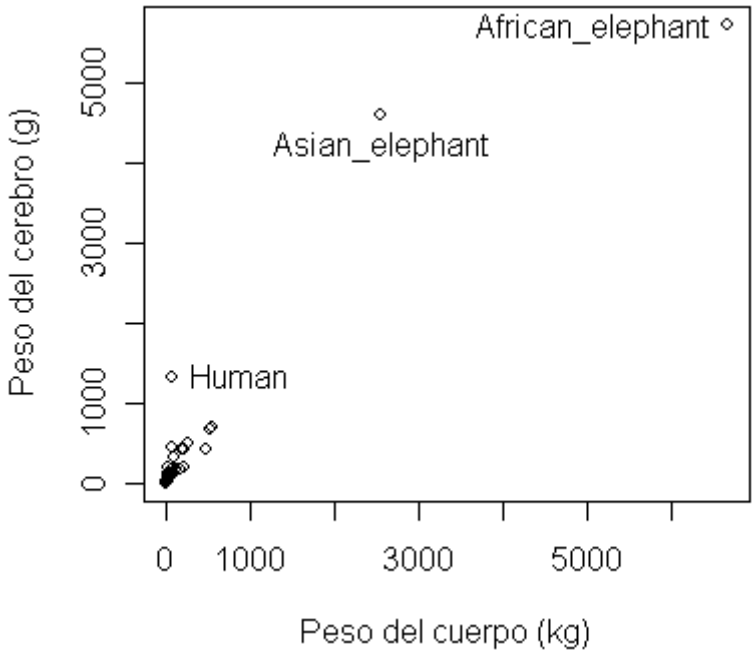

Figura 7.1 Diagrama de dispersión del peso del cerebro versus el peso del cuerpo de 62 especies de mamíferos

7.1.1 Transformaciones potencia

Una *familia de transformaciones* es una colección de transformaciones indexadas por uno o más parámetros que el analista debe seleccionar. La familia más usada es llamada *familia potencia*, definida para una variable estrictamente positiva U por

$$\psi(U,\lambda) = U^{\lambda} \tag{7.1}$$

Conforme el parámetro λ varia, se obtienen los miembros de esta familia, incluyendo las transformaciones raíz cuadrada y cúbica, $\lambda = 1/2$ o $1/3$, e inversa, $\lambda = -1$. Se interpreta el valor de $\lambda = 0$ como la *transformación logarítmica*. Los valores usuales de λ se encuentran en el rango de – 2 a 2, pero por lo general se seleccionan valores entre – 1 y 1. El valor de $\lambda = 1$ *no corresponde a transformación alguna*. La variable U debe ser estrictamente positiva para poder usar estas trasformaciones, sin embargo más adelante se consideran transformaciones para variables que toman valores negativos e incluso el valor cero. Se introduce la notación ψ buscando mayor claridad en la discusión, ya que más adelante se consideran otras familias de transformaciones.

La Figura 7.2 muestra los gráficos de $\psi(\text{PesoCerebro},\lambda)$ versus $\psi(\text{PesoCuerpo},\lambda)$ con el mismo valor de λ para ambas variables, $\lambda = -1, 0, 1/3, 1/2$. No es necesario aplicar la misma transformación a las dos variables, sin embargo en esta situación es razonable ya que ambas variables representan el mismo tipo de medición. Si se permite que cada variable tenga su propia transformación, la búsqueda visual de las transformaciones adecuadas es más difícil ya que se deben considerar muchas más posibilidades.

```
par(mfrow=c(2,2),mai=c(0.6,0.6,0.1,0.1),mgp=c(2,1,0),cex.lab=1.0,cex=0.6)

PesoCuerpo1=1/PesoCuerpo
PesoCerebro1=1/PesoCerebro
m1=lm(PesoCerebro1~PesoCuerpo1)
plot(PesoCuerpo1, PesoCerebro1, xlab = expression(paste("(a)  ",PesoCuerpo^-1)), ylab =
expression(PesoCerebro^-1))
abline(m1)
lines(lowess(PesoCerebro1~PesoCuerpo1,f=2/3,iter=1),lty=2)

PesoCuerpo2=log(PesoCuerpo)
PesoCerebro2=log(PesoCerebro)
m2=lm(PesoCerebro2~PesoCuerpo2)
plot(PesoCuerpo2, PesoCerebro2, xlab = expression(paste("(b)  ",log[e](PesoCuerpo))),
ylab = expression(log[e](PesoCerebro)))
abline(m2)
lines(lowess(PesoCerebro2~PesoCuerpo2,f=2/3,iter=1),lty=2)

PesoCuerpo3=(PesoCuerpo)^(1/3)
PesoCerebro3=(PesoCerebro)^(1/3)
m3=lm(PesoCerebro3~PesoCuerpo3)
plot(PesoCuerpo3, PesoCerebro3, xlab = expression(paste("(c)  ",PesoCuerpo^0.33)), ylab
= expression(PesoCerebro^0.33))
abline(m3)
lines(lowess(PesoCerebro3~PesoCuerpo3,f=2/3,iter=1),lty=2)

PesoCuerpo4=(PesoCuerpo)^(1/2)
PesoCerebro4=(PesoCerebro)^(1/2)
m4=lm(PesoCerebro4~PesoCuerpo4)
plot(PesoCuerpo4, PesoCerebro4, xlab = expression(paste("(d)  ",PesoCuerpo^.5)), ylab =
expression(PesoCerebro^.5))
abline(m4)
lines(lowess(PesoCerebro4~PesoCuerpo4,f=2/3,iter=1),lty=2)
```

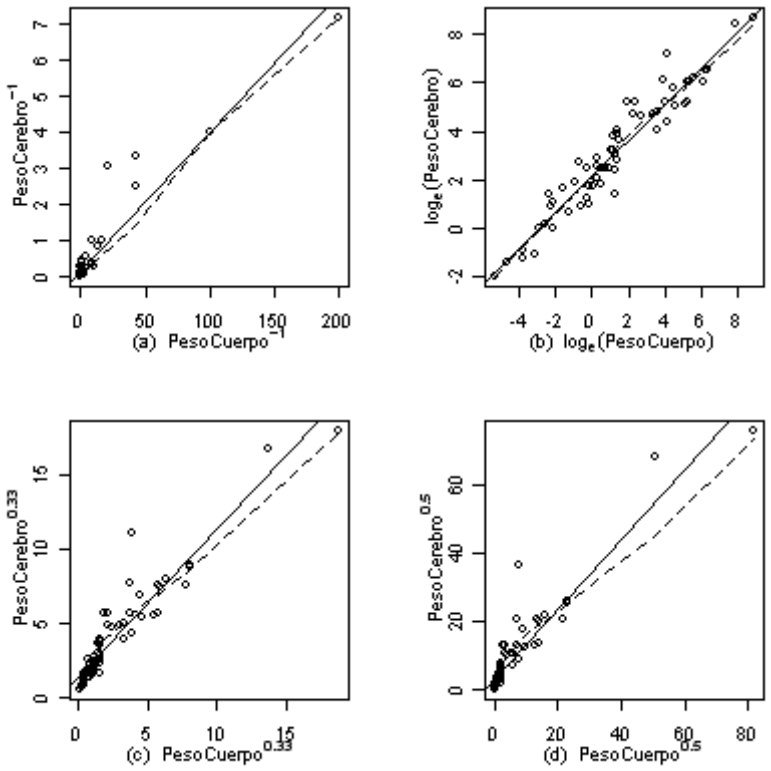

Figura 7.2 Diagramas de dispersión para la data Cerebro con cuatro posibles transformaciones. La línea sólida es la línea por MCO y la línea punteada es el suavizado loess.

A partir de los cuatro gráficos de la Figura 7.2, la elección es reemplazar los pesos por sus logaritmos. En esta escala, la función media parece ser lineal, con el suavizado bastante cerca de la línea por MCO. Adicionalmente, la función variancia en el gráfico de logaritmos parece ser constante.

El uso de logaritmos para la data Mamiferos no es particularmente sorprendente, a la luz de dos reglas empíricas que son bastante útiles en el modelamiento de regresión lineal:

> *La regla de logaritmos* Si el valor del rango de la variable es mayor que uno, en orden de magnitud, y la variable es estrictamente positiva, entonces reemplazar la variable por su logaritmo podría ser útil.
>
> *La regla del rango* Si el rango de una variable es considerablemente menor que uno, en orden de magnitud, entonces cualquier transformación de la variable es poco probable que sea útil.

La regla del logaritmo se satisface para ambos, PesoCuerpo con rango 0.005 kg a 6654 kg y para PesoCerebro, con rango 0.14 g a 5712 g. La transformación logarítmica es un buen punto de partida para examinar cualquier otra transformación sobre las variables.

La regresión lineal simple parece ser apropiada para ambas variables en la escala logarítmica. Lo anterior corresponde al *modelo físico*

$$\text{PesoCerebro} = \alpha \times \text{PesoCuerpo}^{\beta_1} \times \delta \tag{7.2}$$

donde δ es un *error multiplicativo*, tal que el peso promedio del cerebro de una especie en particular se obtienen multiplicando el peso del cuerpo por δ. Se espera que δ tenga media 1 y una distribución concentrada en valores cercanos a 1. Utilizando logaritmos y tomando $\beta_1 = \log\alpha$ y $e = \log\delta$,

$$\log\left(\text{PesoCerebro}\right) = \beta_0 + \beta_1 \log\left(\text{PesoCuerpo}\right) + e$$

que es un modelo de regresión lineal simple. Los científicos que estudian la relación entre atributos de individuos o especies llaman a (7.2) un modelo *alométrico* y el valor de β_1 juega un rol importante en este tipo de estudios. Se enfatiza, sin embargo, que no todas las transformaciones corresponden a modelos físicos interpretables.

7.1.2 Transformando solo la variable predictora

En el ejemplo de la data Mamiferos. Se requiere la transformación de ambas variables para obtener una función media lineal. En otros problemas, solo es necesario trasformar una variable. Si se desea usar una familia de transformaciones potencia es conveniente introducir la familia de *transformaciones potencia escaladas*, definida para X, estrictamente positiva, por

$$\psi_S\left(X,\lambda\right) = \begin{cases} \left(X^{\lambda} - 1\right)/\lambda & \text{si } \lambda \neq 0 \\ \log\left(X\right) & \text{si } \lambda = 0 \end{cases} \tag{7.3}$$

Las transformaciones potencia escaladas $\psi_S\left(X,\lambda\right)$ difieren de las transformaciones potencia $\psi\left(X,\lambda\right)$ en algunos aspectos. Primero $\psi_S\left(X,\lambda\right)$ es una función continua de λ. Como $\lim_{\lambda\to 0}\psi_S\left(X,\lambda\right) = \log_e\left(X\right)$ por lo que la transformación logarítmica es un miembro de esta familia cuando $\lambda = 0$. Además, $\psi_S\left(X,\lambda\right)$ preserva la dirección de la asociación, en el sentido que si $\left(X,Y\right)$ se encuentran relacionadas de manera positiva, entonces $\left(\psi_S\left(X,\lambda\right),Y\right)$ mantiene el mismo tipo de relación para todos los valores de λ. Con las transformaciones potencia, la dirección de la asociación cambia cuando $\lambda < 0$.

Si se consigue una potencia adecuada usando las transformaciones potencia escaladas, en la práctica podría usarse la transformación potencia $\psi\left(X,\lambda\right)$ en el modelamiento de la regresión, ya que ambos difieren solo en escala, locación y posiblemente cambios de signo. Las transformaciones escaladas son usadas solo para seleccionar la transformación a utilizar.

Si se transforma solo el predictor usando una transformación de la familia potencia se tiene la siguiente función media

$$\mathrm{E}[Y/X] = \beta_0 + \beta_1 \psi_S(X, \lambda) \tag{7.4}$$

Si se conoce λ, se puede estimar (7.4) vía MCO y obtener la suma de cuadrados residual, $\mathrm{SCRes}(\lambda)$. El estimado $\hat{\lambda}$ es simplemente el valor de λ que minimiza $\mathrm{SCRes}(\lambda)$. Como regla práctica, no es necesario conocer exactamente λ y por lo general es suficiente seleccionar el valor de

$$\lambda \in \{-1, -1/2, 0, 1/3, 1/2, 1\} \tag{7.5}$$

Como ejemplo, considere la dependencia de la Altura del árbol en decímetros, sobre el diámetro del árbol, Dbh en mm, a 137 cm de una muestra de árboles de cedro en 1991. La data se encuentra en Arboles. La Figura 7.3 muestra el gráfico de dispersión de la data sobre la que se han superpuesto tres curvas. Para cada λ se calculan los valores estimados $\hat{y}(\lambda)$ de la regresión por MCO de Peso sobre $\psi_S(\mathrm{Dbh}, \lambda)$. La línea para un valor particular de λ se obtiene graficando los puntos $(\mathrm{Dbh}, \hat{y}(\lambda))$ uniéndolos con una línea. Entre los tres valores de λ utilizados en la figura, $\lambda = 0$ parece estimar mejor la data. Para $\lambda = 1$ no se logra un buen estimado para árboles grandes y pequeños, mientras que la inversa es demasiado curva para árboles grandes. Lo anterior sugiere reemplazar Dbh por el log(Dbh) tal como se observa en la Figura 7.4.

Como metodología alternativa se puede estimar el parámetro de transformación usando mínimos *cuadrados no lineales*. La función media (7.4) es una función no lineal del parámetro ya que β_1 multiplica la función $\psi_S(X, \lambda)$ no lineal del parámetro λ. Usando los métodos descritos en el Capitulo 11 se obtiene $\hat{\lambda} = 0.05$ con un error estándar de 0.15.

```
Arboles <- read.table(file = "http://tarwi.lamolina.edu.pe/~clopez/Regresion/Arboles.txt", header = T)
attach(Arboles)
library(alr3)
lamda=c(1,0,-1)
nuevo=seq(min(Dbh),max(Dbh),length=100)
plot(Dbh,Altura,cex=0.7)
for (j in 1:3)
    {
       m1=lm(Altura~bcPower(Dbh,lamda[j]))
       lines(nuevo,predict(m1,data.frame(Dbh=nuevo)),lty=j,col=j)
    }
legend(940,200,legend=as.character(lamda),cex=.75,lty=1:3,col=1:3,xjust=1,yjust=1)
```

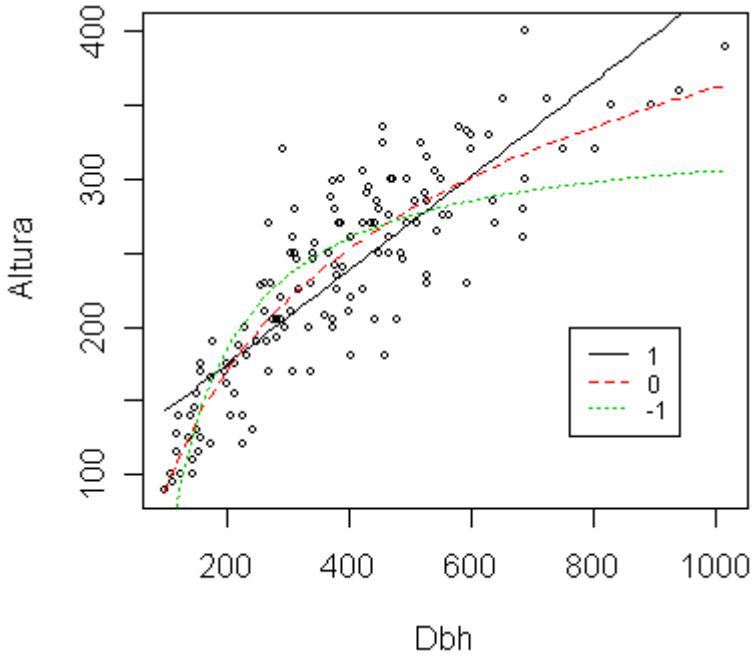

Figura 7.3 Altura versus Dbh para la data Altura.

```
plot(logb(Dbh,2),Altura, ylab="Altura", xlab=expression(log[2](Dbh)))
abline(lm(Altura~logb(Dbh,2)))
```

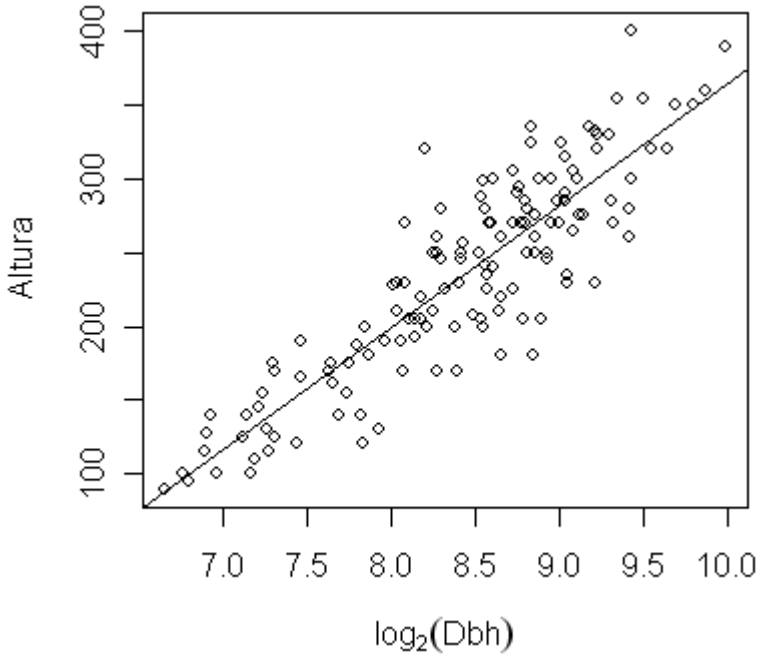

Figura 7.4 La data Altura transformada.

7.1.3 Transformando solo la variable respuesta

Una transformación sobre la variable respuesta puede ser seleccionada usando un *gráfico de valor estimado inverso*, en el cual se toman los valores estimados de la regresión de Y sobre X en el eje vertical y la variable respuesta en el eje horizontal. En la regresión simple los valores estimados son proporcionales al predictor X, luego se trata de un gráfico equivalente para X sobre el eje horizontal y Y sobre el vertical. Es decir, para estimar una transformación $\psi_S\left(Y,\lambda_y\right)$ se trabaja con la función media

$$\mathrm{E}\left[\hat{y}/Y\right]=\alpha_0+\alpha_1\psi_S\left(Y,\lambda_y\right)$$

y se estima λ_y. Un ejemplo del uso del gráfico de variable respuesta inverso se muestra en la Sección 7.3.

7.1.4 El método de Box y Cox

Box y Cox (1964) proporcionan un método general para seleccionar transformaciones de la variable respuesta y que es aplicable tanto a la regresión simple como múltiple. Así como en los métodos previos, se debe seleccionar la transformación a partir de una familia indexada por el parámetro λ. Para el método de Box y Cox se necesita una versión ligeramente más complicada de las familias potencia, llamadas *familia potencia modificada*, definidas por Box y Cox (1964) para una variable respuesta Y, estrictamente positiva, por

$$\begin{aligned}\psi_M\left(Y,\lambda_y\right)&=\psi_S\left(Y,\lambda_y\right)\times\mathrm{gm}\left(Y\right)^{1-\lambda_y}\\ \psi_M\left(Y,\lambda_y\right)&=\begin{cases}\mathrm{gm}\left(Y\right)^{1-\lambda_y}\times\left(Y^{\lambda_y}-1\right)\Big/\lambda_y & \text{si } \lambda_y\neq 0\\ \mathrm{gm}\left(Y\right)\times\log\left(Y\right) & \text{si } \lambda_y=0\end{cases}\end{aligned} \tag{7.6}$$

donde $\mathrm{gm}\left(Y\right)$ es la *media geométrica* de la variable no transformada. Si los valores de Y son $y_1,\cdots,y_n$ entonces la media geométrica de Y es $gm\left(Y\right)=\exp\left(\sum\log\left(y_i\right)/n\right)$.

En el método de Box y Cox se asume que la función media

$$\mathrm{E}\left[\psi_M\left(Y,\lambda_y\right)\Big/\mathbf{X}=\mathbf{x}\right]=\boldsymbol{\beta}'\mathbf{x} \tag{7.7}$$

es correcta para cualquier λ_y. Si λ_y fuese conocido se podría estimar la función media (7.7) usando MCO ya que la variable respuesta transformada $\psi_M\left(Y,\lambda_y\right)$ estaría completamente especificada. Se denota la suma cuadrados residual de esta regresión por $\mathrm{SCRes}\left(\lambda_y\right)$. Se estima λ_y como el valor de la transformación potencia que minimiza $\mathrm{SCRes}\left(\lambda_y\right)$. Desde un punto de vista práctico se puede seleccionar λ_y a partir de (7.5).

El método de Box y Cox no busca linealidad, en realidad se trata de una transformación que busca *normalidad*: se escoge λ de tal forma que los residuales de la regresión de $\psi\left(Y,\lambda_y\right)$ sobre X tenga distribución aproximadamente normal, tanto como sea posible. Hernández y Johnson (1980) recomiendan chequear gráficamente el supuesto anterior antes de seleccionar una transformación. El método de Box y Cox permite encontrar intervalos de confianza para el parámetro de transformación.

7.2 TRANSFORMACIONES Y MATRICES DE DISPERSION

La data Carreteras descrita en la Tabla 7.1 fue tomada de un paper en ingeniería civil de Carl Hoffstedt. En este trabajo se relaciona la tasa de accidentes automovilísticos, en accidentes por millón de millas para vehículos, con varios términos potenciales. La data incluye 39 secciones de carreteras en el estado de Minnesota en 1973. El objetivo de este análisis fue entender el impacto las variables de diseño: Accpts, VLim, Señales y Borde que se encuentran bajo el control del departamento de carreteras, sobre los accidentes. Las restantes variables se encuentran relativamente bajo control del departamento de carreteras y son incluidas para reducir la variabilidad debido a aquellos factores incontrolables. No existe alguna razón particular para creer que la Tasa es una función lineal de los predictores, o alguna razón teórica para preferir alguna forma particular para la función media.

Tabla 7.1 La data Carreteras[a]

Variable	Descripción
Tasa	Tasa de accidentes en 1973 por millón de millas para vehículos
Longitud	Longitud de los segmentos en millas
TDP	Trafico diario promedio estimado (en miles)
Volumen	Volumen del auto como porcentaje del total
VLim	Velocidad limite en 1973
Borde	Ancho del borde externo de la vía (en pies)
Señales	Número de intercambios señalizados por milla en los segmentos

[a] Las variable adicionales de esta data se describen en la Tabla 10.5

Un primer paso, importante, en este análisis es examinar la matriz de dispersión para todos los predictores y la variable respuesta, tal como se muestra en la Figura 7.5. Existen algunas observaciones sobre este gráfico que podrían ayudar a seleccionar las transformaciones:

1. La variable Señales, el número de semáforos por milla, es cero para segmentos de carretera sin semáforo y puede estar por encima de dos para otros segmentos. Las transformaciones pueden ayudar con esta variable, pero como tiene valores no positivos no es posible usar las transformaciones potencia directamente. Como Señales se calcula como el número de señales dividido por Longitud, se reemplaza por Señales1 definido por

$$\text{Señales1} = \frac{\text{Señales} \times \text{Longitud} + 1}{\text{Longitud}}$$

Esta variable siempre es positiva y puede ser transformada usando la familia potencia.

2. TDP y Longitud tienen un rango alto, probablemente los logaritmos puedan resultar apropiados para ellos.
3. VLim varía de 40 a 70 kilómetros por hora, por lo general entre 50 y 60. Las transformaciones quizás no sean de mucha utilidad.
4. Cada uno de los predictores se encuentran modestamente asociados con Tasa, tal como lo muestran los gráficos en la parte superior de la Figura 7.5.
5. Muchos de los predictores se encuentran relacionados. En algunos casos, la función media entre predictores es bastante lineal y en otros no tanto.

```
Carreteras <- read.table(file = "F:Carreteras.txt", header = T)
attach(Carreteras)
pairs(Tasa ~ Longitud + TDP + Volumen + VLim + Borde + Señales)
```

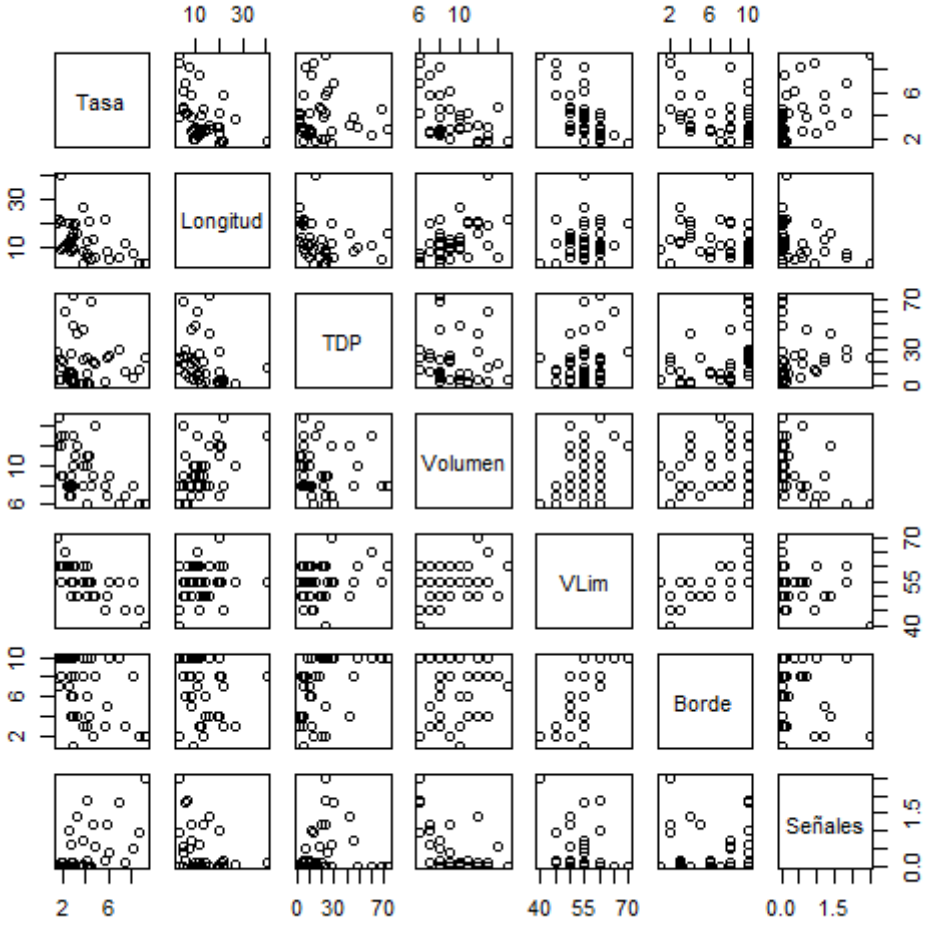

Figura 7.5 La data Carreteras no transformada.

Luego de este análisis preliminar de la matriz de dispersión se tiene la tarea de encontrar una buena transformación. Lo anterior trae la siguiente pregunta: ¿Cuál es el objetivo de seleccionar una transformación? ¿Cómo decidir si se hizo una buena elección?

El objetivo principal de seleccionar una transformación en un problema de regresión lineal simple, es encontrar aquella en la que el modelo aproxime bastante bien la data. La conexión entre el objetivo mencionado y la selección de una transformación no es completamente obvia. Los trabajos de Brillinger (1983) y Li y Duan (1989) proporcionan una conexión teórica. Suponga que se tiene una variable respuesta Y y un conjunto de predictores $\mathbf{X}$, y suponga que se cumple

$$\mathrm{E}\left[Y/\mathbf{X}=\mathbf{x}\right]=g\left(\boldsymbol{\beta}'\mathbf{x}\right) \tag{7.8}$$

para alguna *función no especificada y completamente desconocida*. De acuerdo a lo anterior, la media de Y depende de $\mathbf{X}$ a través de una combinación lineal de los términos en $\mathbf{X}$, y el gráfico de Y versus $\boldsymbol{\beta}'\mathbf{x}$ debería mostrar a g y su función media. Luego, se podría estimar g o transformar Y para convertir la función media en lineal.

7.2.1 Selección automática de las transformaciones para los predictores

Usando los resultados de la Sección 7.2.1, se buscaba una transformación de los predictores tal que los gráficos de dispersión de un predictor versus otro tengan una función media lineal, o que al menos no sea demasiada curva. Sin herramientas gráficas interactivas o algún método automático de selección de las transformaciones, lo anterior podría ser una tarea desalentadora ya que el analista necesitaría obtener muchas matrices de dispersión hasta obtener un conjunto útil de transformaciones.

Velilla (1993) propuso una extensión multivariada del método de Box-Cox para seleccionar las transformaciones que proporcionan linealidad, y a la vez éste método proporciona un muy buen punto de partida para seleccionar las transformaciones. Para un conjunto de k predictores no transformados y estrictamente positivos $\mathbf{X}=\left(X_1 \quad \cdots \quad X_k\right)$ se aplicara una transformación potencia modificada a cada X_j, el conjunto de los k parámetros de transformación se encuentran en $\boldsymbol{\lambda}=\left(\lambda_1 \quad \cdots \quad \lambda_k\right)'$. Sea $\psi_M\left(\mathbf{X},\boldsymbol{\lambda}\right)$ el conjunto de variables

$$\psi_M\left(\mathbf{X},\boldsymbol{\lambda}\right)=\left(\psi_M\left(X_1,\lambda_1\right) \quad \cdots \quad \psi_M\left(X_k,\lambda_k\right)\right)$$

Sea $\mathbf{V}\left(\boldsymbol{\lambda}\right)$ la matriz de covariancias muestrales de la data transformada $\psi_M\left(\mathbf{X},\boldsymbol{\lambda}\right)$. El valor $\hat{\boldsymbol{\lambda}}$ se selecciona como el valor de $\boldsymbol{\lambda}$ que minimiza el logaritmo del determinante de $\mathbf{V}\left(\boldsymbol{\lambda}\right)$. Este proceso de minimización puede ser llevado a cabo usando algún lenguaje de alto nivel como R, S plus, Excel, etc. El proceso de minimización requiere de la especificación de la función a ser minimizada y los valores iniciales del algoritmo. Estos valores se toman como $\boldsymbol{\lambda}=\mathbf{0}$ o $\boldsymbol{\lambda}=\mathbf{1}$ o algún vector apropiado de ceros y unos.

Retornado a la data Carretera, a la cual se elimino la variable VLim debido a que su rango es bastante estrecho. Para los términos restantes se obtiene, en la Tabla 7.2, el resumen de las transformaciones usando el método multivariado de Box-Cox. La tabla muestra los valores de $\hat{\boldsymbol{\lambda}}$ en la columna "Est. Power". Para nuestros propósitos, los errores estándar pueden tratarse como si estuvieran asociados a los coeficientes de regresión. Las siguientes dos columnas sirven para verificar si el parámetro de transformación es igual a cero o uno. Los valores obtenidos pueden ser comparados con la distribución normal, de tal forma que los valores mayores a 1.96 en valor absoluto corresponden a p-valores menores que 0.05. Las potencias para Longitud, TDP, Borde y Señales1 no parecen ser diferentes de cero y para Borde no aparenta ser diferente de uno. Al final de la tabla se encuentran las *pruebas de razón de verosimilitud*. La primera sirve para probar que todas las potencias son cero lo cual es firmemente rechazado ya que el valor aproximado $\chi^2(5)$ es bastante grande. Similarmente, la segunda que sirve para probar que no es necesario realizar transformaciones $(\boldsymbol{\lambda}=\mathbf{1})$ es rechazada. La última fila de resultados sirve para probar que las tres primeras variables deberían estar en escala logarítmica y la última no transformada, con un p-valor de 0.29. Los predictores en la escala transformada junto a la respuesta se muestran en la Figura 7.6. Todos los gráficos 2D tiene función media lineal, al menos aproximadamente. Lo anterior proporciona un buen punto de partida para el modelamiento de regresión.

```
Señales1 <- (Señales*Longitud+1)/Longitud
Transf <- powerTransform(cbind(Longitud, TDP, Volumen, Borde, Señales1) ~ 1)
summary(Transf)
```

```
bcPower Transformations to Multinormality

          Est.Power Std.Err. Wald Lower Bound Wald Upper Bound
Longitud     0.1437   0.2127          -0.2732           0.5607
TDP          0.0509   0.1206          -0.1854           0.2872
Volumen     -0.7028   0.6177          -1.9134           0.5078
Borde        1.3456   0.3630           0.6341           2.0570
Señales1    -0.2408   0.1496          -0.5341           0.0525

Likelihood ratio tests about transformation parameters
                                     LRT df         pval
LR test, lambda = (0 0 0 0 0)  23.324467  5 0.0002926014
LR test, lambda = (1 1 1 1 1) 132.857421  5 0.0000000000
LR test, lambda = (0 0 0 1 0)   6.088599  5 0.2976930877
```

```
plot(Transf)
```

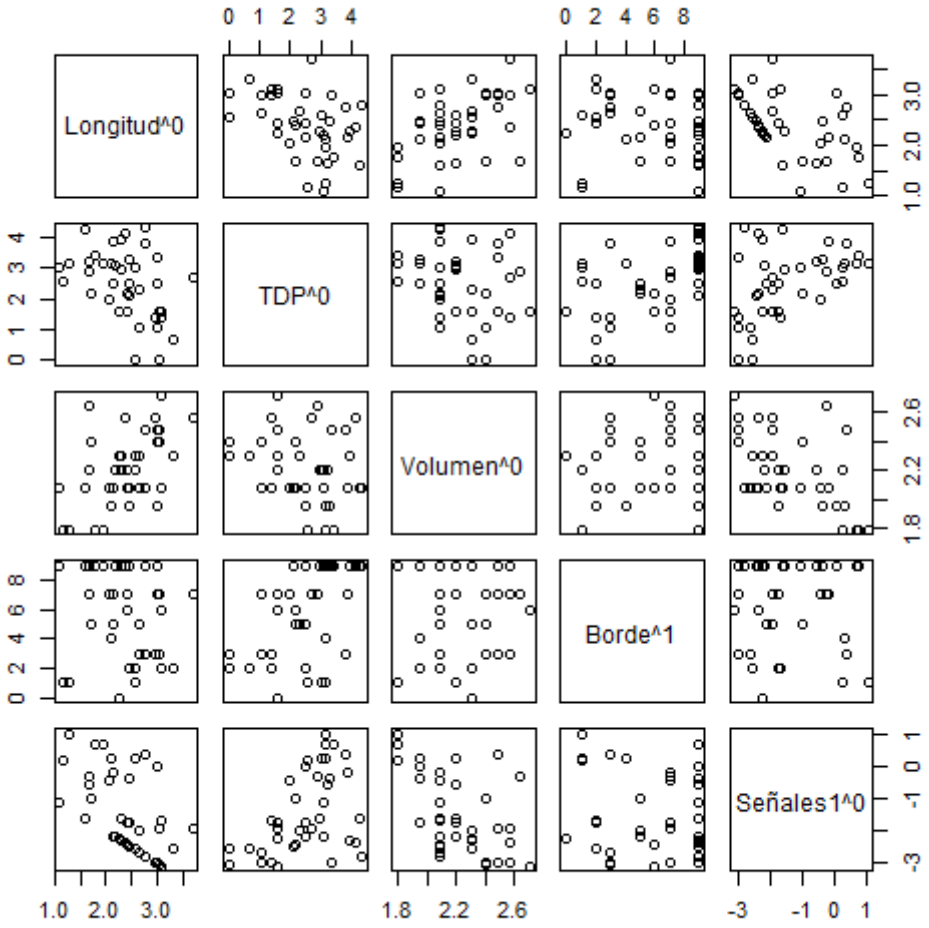

Figura 7.6 Predictores transformados para la data Carretera.

7.3 TRANSFORMANDO LA VARIABLE RESPUESTA

Una vez transformados los términos, es turno de transformar la variable respuesta. La Figura 7.7 es el gráfico de valor estimado inverso para la data Carretera usando los términos transformados determinados en la sección anterior. En este gráfico la variable respuesta Tasa se encuentra en el eje horizontal y los valores estimados de la regresión sobre los términos transformados en el eje vertical. Cook y Weisberg (1994) mostraron que si los predictores tienen relación lineal de manera aproximada entonces pueden usarse los métodos de la Sección 7.1.2 para seleccionar una transformación para la variable respuesta. Entre las tres curvas mostradas en este gráfico, la transformación logarítmica parece ser la más apropiada.

El método de Box-Cox proporciona un procedimiento alternativo para encontrar una transformación de la variable respuesta. Este método se resume mediante un gráfico con λ_y en el eje horizontal y $\text{SCRes}(\lambda_y)$ o mejor aún $-(n/2)\log(\text{SCRes}(\lambda_y)/n)$ sobre el eje vertical. Si se elige el último término, el estimado $\hat{\lambda}_y$ es el punto que maximiza la curva. Un intervalo de confianza esta dado por el conjunto de todos los λ_y tal que $\log(\text{L}(\hat{\lambda}_y)) - \log(\text{L}(\lambda_y)) < 1.92$. Este gráfico para la data Carretera se muestra en la Figura 7.8, con $\hat{\lambda} \approx -0.2$ y un intervalo de confianza que va de – 0.8 a + 0.3. La transformación logarítmica se encuentra en el intervalo de confianza, de acuerdo con el gráfico inverso de valor estimado.

Para la data Carretera, los dos métodos de transformación de la variable respuesta parecen estar de acuerdo, pero no existe razón teórica para que esto necesariamente suceda. Los siguientes pasos se recomiendan para seleccionar una transformación para la variable respuesta.

1. Con predictores aproximadamente lineales, construir el gráfico inverso de $\hat{y}$ versus la variable respuesta. Si éste muestra una clara tendencia no lineal, entonces la variable respuesta debe ser transformada para ajustar esta tendencia. No existe razón por la que solo deberían ser consideradas las transformaciones potencia. Por ejemplo, la transformación podría ser seleccionada usando un suavizador. Si no existe una clara tendencia no lineal, las transformaciones para la variable respuesta están lejos de ser útiles.
2. El procedimiento de Box-Cox puede ser usado para seleccionar una transformación que proporcione normalidad. Lo anterior requiere el uso de una familia de transformaciones.

Para la data Carretera, se tiene un punto de partida razonable para el análisis de regresión, con muchos de los predictores y la variable respuesta transformados a la escala logarítmica. Se retomara este ejemplo en los capítulos finales.

```
Carretera1 <- cbind(Tasa = Tasa, logLong = log2(Longitud), logTDP = log2(TDP), logVol = log2(Volumen), VLim, Borde, logSeñ = log2(Señales1))
m2 <- lm(Tasa ~ logLong + logTDP + logVol + VLim + Borde + logSeñ, data = data.frame(Carretera1))
invTranPlot(Tasa, predict(m2), ylab = "Valores estimados")
```

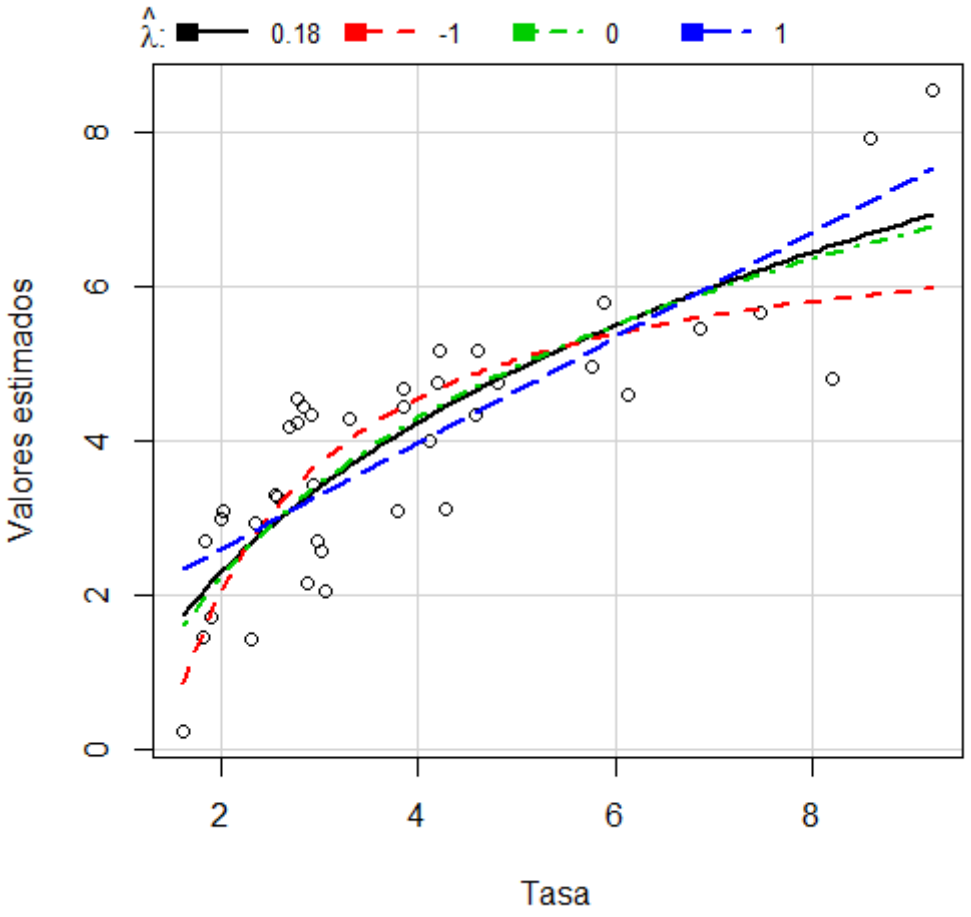

Figura 7.7 Gráfico inverso de valor estimado para la data Carretera.

```
      lambda      RSS
1  0.1846179 30.62130
2 -1.0000000 34.71600
3  0.0000000 30.72613
4  1.0000000 32.45792
```

boxCox(m2, xlab=expression(lambda[y]))

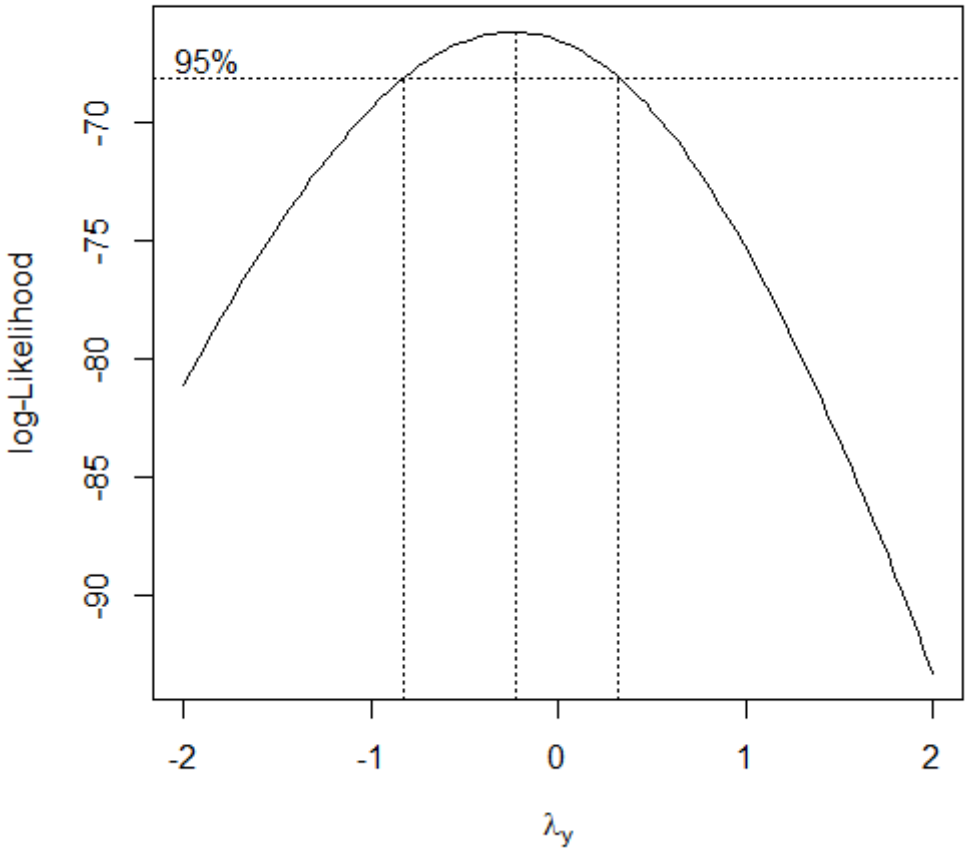

Figura 7.8 Resumen gráfico Box-Cox para la data Carretera.

7.4 TRANSFORMACIONES PARA VARIABLES NO POSITIVAS

Se han sugerido familias de transformaciones para una variable U que incluye valores negativos. La idea central es usar los métodos discutidos en este capítulo para seleccionar una transformación que permita que U tome valores negativos. Una posibilidad es considerar transformaciones de la forma $(U+\gamma)^{\lambda}$, donde γ es lo suficientemente grande para asegurar que $U+\gamma$ es estrictamente positivo. Se usa una variante de este método con la variable Señales en la data Carretera. En principio, (γ,λ) puede ser estimado simultáneamente, aunque en la práctica los estimados de γ son muy variables y poco confiables. Alternativamente, Yeo y Jonson (2000) propusieron una familia de transformaciones que pueden ser usadas sin restricciones sobre U y tienen muchas de las buenas propiedades de la familia potencia de Box-Cox. Estas transformaciones están definidas por

$$\psi_{YJ}(U,\lambda)=\begin{cases}\psi_M(U+1,\lambda) & \text{si } U\geq 0\\ M(-U+1,2-\lambda) & \text{si } U<0\end{cases} \tag{7.10}$$

Si U es estrictamente positiva, entonces la transformación de Yeo-Johnson es la misma que la transformación potencia de Box-Cox para $(U+1)$. Si U es estrictamente negativa, entonces la transformación de Yeo-Johnson es la transformación potencia de Box-Cox para $(-U+1)$, pero con potencia $2-\lambda$. Con valores positivos y negativos, la transformación es una mixtura de los dos, es decir potencias diferentes para valores positivos y negativos. En este último caso, la interpretación del parámetro de transformación es difícil, ya que tiene diferentes significados para $U \geq 0$ y para $U < 0$. La Figura 7.9 muestra las transformación de Box-Cox y Yeo-Johnson para los valores de $\lambda = -1, 0, 0.5$. Para valores positivos, las dos transformaciones difieren en su comportamiento con valores cercanos a cero, con la transformación de Box-Cox proporcionando un cambio mucho mayor para pequeños valores en comparación con Yeo-Johnson.

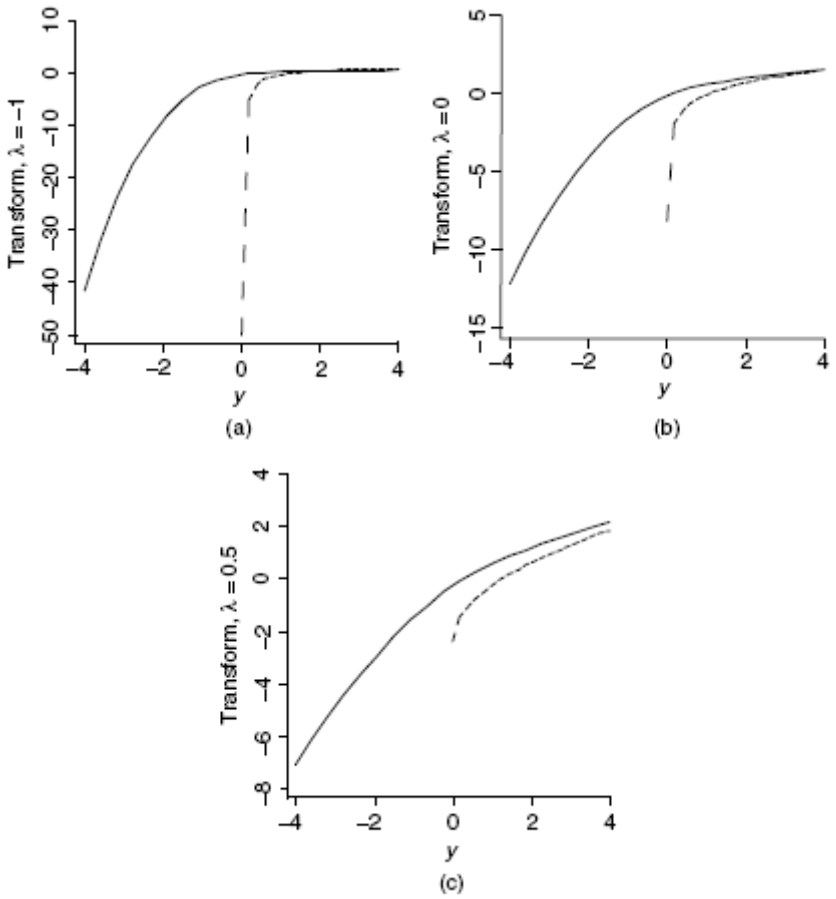

Figura 7.9 Comparación de las transformaciones de Box-Cox y Yeo-Johnson.