

**Modelo de Regresión para razones y
proporciones:
Variantes y aplicaciones
socio-económicas y biosanitarias**



Trabajo Fin de Grado

Dept. de Estadística e Investigación Operativa

FACULTAD DE MATEMÁTICAS

Curso 2014-2015

Autor: Ana María Real Pérez

Tutor: Juan Manuel Muñoz Pichardo

Abstract

Many researchers use regression models to analyse data and study the relationship between some variables. The linear regression model, in particular, is commonly used in applications. This paper proposes a model based on the assumption that the response is beta distributed, which is very flexible for modelling proportions. For this reason, this model is appropriate for situations where the response is restricted to the interval $(0, 1)$. The goal is to collect the modelling and inferential procedures for beta regression, including a description of `betareg` package which provides the class of beta regressions in the R system for statistical computing. Furthermore, some practical applications that employ real data are mentioned and described at the end.

Índice general

1. Distribución beta	7
1.1. Introducción	7
1.2. Caracterización	10
1.3. Propiedades y relaciones	12
1.4. Interpretación de parámetros. Otras parametrizaciones	13
1.5. Aplicaciones	17
2. Regresión beta	19
2.1. El modelo	19
2.2. Estimación de parámetros	21
2.3. Contrastes de hipótesis	25
2.4. Medidas de ajuste	27
2.5. Otras variantes del modelo	28
3. Aplicaciones con R	29
3.1. El paquete <code>Betareg</code>	29
3.2. Aplicaciones prácticas	38

Capítulo 1

Distribución beta

1.1. Introducción

La distribución beta es una familia de distribuciones de probabilidad continuas definidas en el intervalo $[0, 1]$ parametrizadas por dos parámetros, denotados α y β , que controlan la distribución de la variable aleatoria.

La distribución beta es usada para modelar el comportamiento de variables aleatorias limitadas por intervalos de longitud finita. En particular, es una distribución adecuada para porcentajes y proporciones.

Un estudio realizado por Robert Kieschnick, de la Universidad de Texas, y BD McCullough, de la Universidad Drexel [1], compara siete modelos de regresión básicos para proporciones observadas en el intervalo abierto $(0,1)$. Estos modelos son:

1. Distribución normal: función respuesta lineal

En este modelo se asume que la distribución condicionada es una distribución normal y además la esperanza condicionada es una función lineal de unas variables predictoras \mathbf{x} , es decir,

$$f(y|\mathbf{x}) \sim N(k(\mathbf{x}), \sigma^2) \text{ con } k(\mathbf{x}) = \mathbf{x}'\beta$$

2. Distribución normal logística aditiva

Usando la transformación logit en la variable dependiente, $\ln(y/(1 - y))$, tenemos el modelo

$$\ln\left(\frac{y}{1-y}\right) = \mathbf{x}'\beta + \epsilon$$

donde se asume que ϵ se distribuye según una normal estándar. Si la variable y sigue una distribución normal logística aditiva, se puede demostrar que la variable $z = \ln(y/(1 - y))$ sigue una distribución normal.

3. Distribución normal censurada

Modelo en el cual definimos la variable

$$y_i = \begin{cases} 0, & y_i^* \leq 0 \\ y_i^*, & 0 < y_i^* < 1 \\ 1, & y_i^* \geq 1 \end{cases}$$

asumiendo que y_i^* se distribuye según una normal y que se puede escribir como

$$y_i^* = \mathbf{x}'_i \beta + u_i, \quad i = 1, 2, \dots, n$$

donde u_i se consideran independientes e idénticamente distribuidas según una normal estándar.

4. Distribución normal: función respuesta no lineal

Para el estudio de la función esperanza condicionada muchos investigadores recomiendan usar la función logística acumulada, que viene dada por:

$$y_i = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} + \epsilon_i$$

donde ϵ_i sigue una distribución normal estándar.

5. Distribución beta

Consideramos que la variable y está distribuida de la siguiente forma:

$$f(y) = \frac{1}{B(p, q)} y^{p-1} (1 - y)^{q-1}$$

donde $0 \leq y \leq 1$ y $B(p, q)$ es la función beta.

Para el modelo de regresión beta podemos seguir el modelo de regresión lineal y asumir que la media es una función lineal de otras variables independientes \mathbf{x} . Es decir, asumimos que la variable y sigue una distribución beta y que

$$E(y|\mathbf{x}) = \frac{p}{p + q} = \mathbf{x}' \beta$$

Otro modelo sería usar la función vínculo logit, que consiste en considerar

$$E(y_i|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{x}_i'\beta)}$$

6. Distribución simple

Otro modelo de regresión para proporciones se basa en la distribución simple, que viene dada por

$$f(y; \mu, \sigma^2) = [2\pi\sigma^2 \{y(1-y)\}^3]^{-1/2} \exp\left\{-\frac{1}{2} d(y; \mu)\right\}$$

para $0 < y < 1$, donde

$$d(y; \mu) = \frac{(y - \mu)^2}{y(1-y)\mu^2(1-\mu)^2}$$

es la desviación, y $0 < \mu < 1$.

El modelo lineal generalizado puede ser aplicado a modelos basados en la distribución simple, usando como función vínculo la función logit.

7. Modelo de regresión quasi-paramétrico

Todos los casos anteriores eran modelos de regresión paramétricos, por lo que el último modelo es el modelo de regresión de quasi-verosimilitud para proporciones del intervalo $(0, 1)$ usando un número finito de observaciones límite, es decir, 0s y 1s:

$$l_i(\beta) = y_i \ln[G(\mathbf{x}_i)] + (1 - y_i) \ln[1 - G(\mathbf{x}_i)]$$

con $G(\cdot)$ una función de las \mathbf{x}_i que verifique $0 < G(\cdot) < 1$. En el estudio se utiliza la función logística en los casos prácticos.

Una vez presentados los modelos de regresión más usados para el análisis de proporciones, éstos son aplicados a datos reales para poder comparar las ventajas y desventajas de cada modelo.

A partir de este análisis comparativo, Kieschnick y McCullough concluyen su estudio recomendando a los investigadores usar en el futuro un modelo de regresión paramétrico basado en la distribución beta.

Este capítulo se estructura en 5 secciones. En la segunda sección se presentan distintas caracterizaciones de la distribución beta, en la sección 1.3 se recogen algunas propiedades de interés de la distribución, en la siguiente

se interpretan los parámetros y se finaliza el capítulo con ilustraciones de las aplicaciones de esta distribución.

Para el desarrollo de este capítulo la bibliografía básica utilizada es:

- JOHNSON Y OTROS (1995). “Chapter 21: Beta Distributions”[2]
- WIKIPEDIA CONTRIBUTORS (2015). “Beta distribution”[3]

1.2. Caracterización

Una variable aleatoria X se dice que sigue una distribución beta de parámetros $\alpha > 0$ y $\beta > 0$, $X \sim Beta(\alpha, \beta)$, si la función de densidad tiene la siguiente forma:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

donde $\Gamma(z)$ es la función gamma y $B(\alpha, \beta)$ es la función beta:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Función de distribución

$$F(x; \alpha, \beta) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)} = I_x(\alpha, \beta)$$

donde $B(x; \alpha, \beta)$ es la función beta incompleta, una generalización de la función beta que viene dada por esta expresión:

$$B(x; \alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

La función $I_x(\alpha, \beta)$ se denomina función beta incompleta regularizada.

Función generatriz de momentos

$$M_X(\alpha; \beta; t) = E[e^{tX}] = 1 + \sum_{i=1}^{\infty} \left(\prod_{r=0}^{i-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^i}{i!}$$

a partir de la cual se obtiene, en general, el momento de orden k :

$$E[X^k] = \prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r}$$

Por lo tanto, el valor esperado y la varianza son:

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Momentos para transformaciones de la variable aleatoria

La esperanza de algunas transformaciones de interés de la variable X son:

- $E[1 - X] = \frac{\beta}{\alpha + \beta}$
- $E[X(1 - X)] = E[(1 - X)X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
- $E\left[\frac{1}{X}\right] = \frac{\alpha + \beta - 1}{\alpha - 1}$ condicionado a $\alpha > 1$
- $E\left[\frac{1}{1 - X}\right] = \frac{\alpha + \beta - 1}{\beta - 1}$ condicionado a $\beta > 1$
- $E\left[\frac{X}{(1 - X)}\right] = \frac{\alpha}{\beta - 1}$ condicionado a $\beta > 1$
- $E\left[\frac{(1 - X)}{X}\right] = \frac{\alpha}{\beta - 1}$ condicionado a $\beta > 1$
- $E[\ln(X)] = \psi(\alpha) - \psi(\alpha + \beta)$, donde $\psi(\cdot)$ es la función digamma definida como la derivada logarítmica de la función gamma:

$$\psi(z) = \frac{d \ln \Gamma}{dz}(z) = \frac{\Gamma'(z)}{\Gamma(z)}$$

1.3. Propiedades y relaciones

Sea X una variable de distribución beta con $\alpha, \beta > 1$. Además de la esperanza y la varianza expresadas anteriormente, se recogen a continuación otras medidas de tendencia central y de dispersión.

Otras medidas de tendencia central.

- Moda

$$\frac{\alpha - 1}{\alpha + \beta - 2}$$

- Media Geométrica

El logaritmo de la media geométrica G_X viene dado por

$$\ln G_X = E[\ln X] = \psi(\alpha) - \psi(\alpha + \beta)$$

y por tanto

$$G_X = e^{\psi(\alpha) - \psi(\alpha + \beta)}$$

- Media Armónica

$$H_X = E\left[\frac{1}{X}\right]$$

Otras medidas de dispersión y forma.

- Desviación media absoluta

Para la distribución beta con parámetros α y β la desviación media absoluta es

$$E[|X - E[X]|] = \frac{2\alpha^\alpha \beta^\beta}{B(\alpha, \beta)(\alpha + \beta)^{\alpha + \beta + 1}}$$

- Asimetría

La asimetría de la distribución beta viene dada por la expresión

$$\gamma_1 = \frac{E[(X - \mu)^3]}{(var(X))^{3/2}} = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}}$$

- Curtosis y coeficiente de curtosis

$$Curtosis = \frac{E[(X - \mu)^4]}{(var(X))^2}$$

$$\gamma_2 = curtosis - 3 = \frac{6[\alpha^3 - \alpha^2(2\beta - 1) + \beta^2(\beta + 1) - 2\alpha\beta(\beta + 2)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}$$

Relaciones entre medidas estadísticas

- Relación entre la moda, la media y la mediana

Expresando la moda y la media en términos de α y β , para $\beta > \alpha > 1$, se tiene la siguiente desigualdad

$$\frac{\alpha - 1}{\alpha + \beta - 2} \leq \text{mediana} \leq \frac{\alpha}{\alpha + \beta}$$

- Curtosis limitada por el cuadrado del coeficiente de asimetría

$$\gamma_1^2 + 1 < curtosis < \frac{3}{2}\gamma_1^2 + 3$$

o, equivalentemente,

$$\gamma_1^2 - 2 < \gamma_2 < \frac{3}{2}\gamma_1^2$$

1.4. Interpretación de parámetros. Otras parametrizaciones

Distribución beta simétrica ($\alpha = \beta$)

En el caso de igualdad de ambos parámetros la función de densidad es simétrica respecto del punto $1/2$, que coincide con el valor de la mediana y la media, y el coeficiente de asimetría vale $\gamma_1 = 0$. En particular podemos diferenciar estos casos:

- $\alpha = \beta < 1$

- Distribución con forma de U
 - Bimodal en 0 y 1, y con un mínimo en $1/2$ (antimoda)
 - $1/12 < \text{var}(X) < 1/4$
 - $-2 < \gamma_2 < 1/4$
 - Si $\alpha = \beta = 1/2$ es la distribución arcoseno.
 - Si $\alpha = \beta \rightarrow 0$ equivale a una distribución Bernoulli con misma probabilidad $1/2$ en $x = 0$ y $x = 1$ y probabilidad cero en cualquier otro punto.
- $\alpha = \beta = 1$
 - Distribución uniforme en $[0, 1]$
 - No hay moda
 - $\text{var}(X) = 1/12$
 - $\gamma_2 = -6/5$
 - $\alpha = \beta > 1$
 - Distribución simétrica unimodal
 - moda= $1/2$
 - $0 < \text{var}(X) < 1/12$
 - $-6/5 \leq \gamma_2 \leq 0$
 - Si $\alpha = \beta = 3/2$ es una distribución semielíptica en $[0, 1]$
 - Si $\alpha = \beta = 2$ es la distribución parabólica en $[0, 1]$
 - Si $\alpha = \beta > 2$ es una distribución con forma de campana
 - Si $\alpha = \beta \rightarrow \infty$ es una distribución donde toda la probabilidad se concentra en el punto $x = 1/2$

Distribución beta asimétrica ($\alpha \neq \beta$)

Cuando los parámetros toman valores distintos la función de densidad deja de ser simétrica. Un intercambio de los valores de los parámetros da lugar a la imagen simétrica de la curva original. Algunos casos particulares son:

- $\alpha < 1, \beta < 1$
 - Distribución con forma de U
 - Coeficiente de asimetría positivo para $\alpha < \beta$ y negativo para $\alpha > \beta$

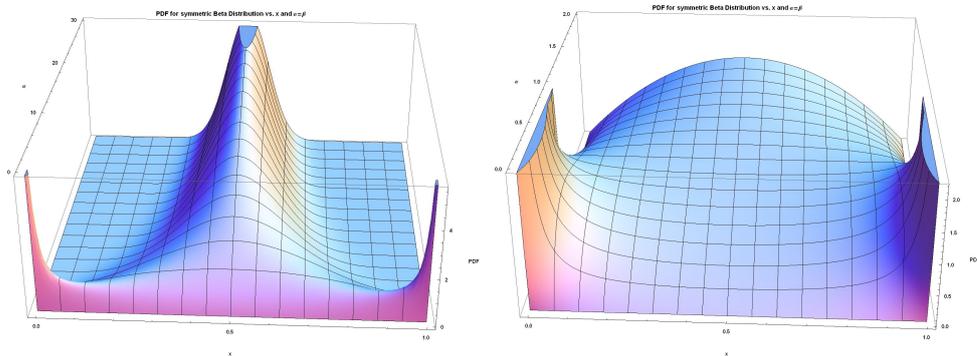


Figura 1.1: Distribución beta con $\alpha = \beta$. En la imagen de la izquierda los parámetros varían de 0 a 30 y en la de la derecha lo hacen de 0 a 2

- Bimodal en 0 y 1, con antimoda en $\frac{\alpha-1}{\alpha+\beta-2}$
 - $0 < \text{mediana} < 1$
 - $0 < \text{var}(X) < 1/4$
- $\alpha > 1, \beta > 1$
 - Unimodal, $\text{moda} = \frac{\alpha-1}{\alpha+\beta-2}$
 - Coeficiente de asimetría positivo para $\alpha < \beta$ y negativo para $\alpha > \beta$
 - $0 < \text{mediana} < 1$
 - $0 < \text{var}(X) < 1/12$
- $\alpha < 1, \beta \geq 1$
 - Forma de J con cola a la derecha
 - Coeficiente de asimetría positivo
 - Distribución estrictamente decreciente y convexa
 - $\text{moda} = 0$
 - $0 < \text{mediana} < 1/2$
 - $0 < \text{var}(X) < \frac{-11+5\sqrt{5}}{2}$

- $\alpha \geq 1, \beta < 1$
 - Forma de J con cola a la izquierda
 - Coeficiente de asimetría negativo
 - Distribución estrictamente creciente y convexa
 - $moda = 1$
 - $1/2 < mediana < 1$
 - $0 < var(X) < \frac{-11+5\sqrt{5}}{2}$
- $\alpha = 1, \beta > 1$
 - Coeficiente de asimetría positivo
 - Distribución estrictamente decreciente
 - Simétrica a la distribución función potencia, $f(x) = x^k$ en el intervalo $[0, 1]$
 - $moda = 0$
 - Si $\alpha = 1$ y $1 < \beta < 2$ la función es cóncava con $1 - 1/\sqrt{2} < mediana < 1/2$ y $1/18 < var(X) < 1/12$
 - Si $\alpha = 1$ y $\beta = 2$ es una recta de pendiente -2 con $mediana = 1 - 1/\sqrt{2}$ y $var(X) = 1/18$
 - Si $\alpha = 1$ y $\beta > 2$ la distribución tiene forma de J con cola a la derecha, es convexa con $0 < mediana < 1 - 1/\sqrt{2}$ y $0 < var(X) < 1/18$
- $\alpha > 1, \beta = 1$
 - Coeficiente de asimetría negativo
 - Distribución estrictamente creciente igual a la distribución función potencia en $[0, 1]$
 - $moda = 1$
 - Si $1 \ll 2$ y $\beta = 1$ la función es cóncava con $1/2 < mediana < 1/\sqrt{2}$ y $1/18 < var(X) < 1/12$
 - Si $\alpha = 2$ y $\beta = 1$ es una recta de pendiente $+2$ con $mediana = 1/\sqrt{2}$ y $var(X) = 1/18$
 - Si $\alpha > 2$ y $\beta = 1$ obtiene una forma de J con cola a la izquierda y convexa con $1/\sqrt{2} < mediana < 1$ y $0 < var(X) < 1/18$

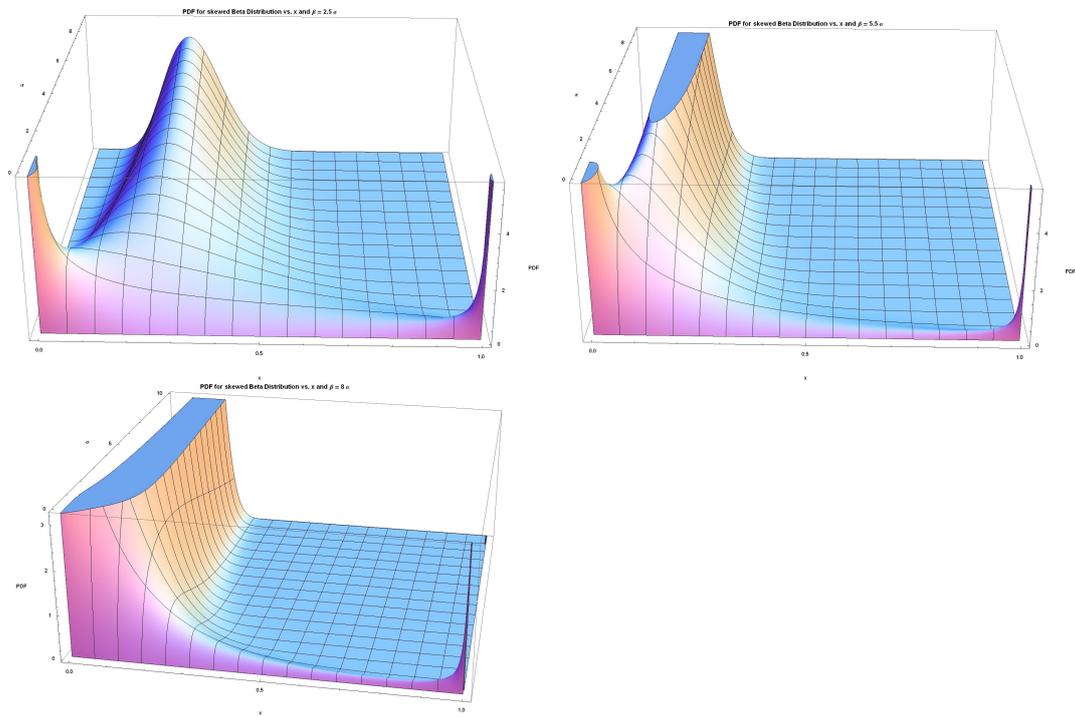


Figura 1.2: **Figura superior izquierda:** distribución beta con $\beta = 2,5$ y α variando de 0 a 9. **Figura superior derecha:** distribución beta con $\beta = 5,5$ y α variando de 0 a 9. **Figura inferior:** distribución beta con $\beta = 8$ y α variando de 0 a 10.

1.5. Aplicaciones

La distribución beta tiene importantes aplicaciones como en la teoría de los estadísticos ordenados. El k -ésimo estadístico ordenado de una muestra de tamaño n de una distribución continua uniforme sigue una distribución beta, es decir, $X_{(k)} \sim Beta(k, n + 1 - k)$.

Esta distribución también tiene uso en la inferencia Bayesiana, ya que produce una familia de distribuciones de probabilidad a priori conjugadas, como por ejemplo $Beta(1, 1)$ (distribución uniforme en el intervalo $(0,1)$), $Beta(0, 0)$, (distribución Bernoulli cuando $\alpha, \beta \rightarrow 0$) y $Beta(1/2, 1/2)$.

Finalmente, como se recoge en la introducción de este capítulo y se ilustra en el capítulo siguiente, la distribución beta juega un papel fundamental en la modelización de razones y proporciones.

Capítulo 2

Regresión beta

2.1. El modelo

El modelo de regresión comúnmente tiene por objetivo modelar y analizar la relación de una variable objetivo con otras variables explicativas a través de un conjunto de datos observados. En particular, el modelo de regresión lineal se usa con frecuencia, sin embargo no es apropiado para situaciones en la que la variable respuesta está restringida al intervalo $(0,1)$. Una posible solución es transformar la variable dependiente para que tome valores en la recta real, y entonces modelar la media de la variable respuesta transformada. Esta aproximación tiene algunos inconvenientes como el hecho de que los parámetros del modelo no pueden ser fácilmente interpretados en términos de la variable original. El objetivo es proponer un modelo de regresión apto para situaciones donde la variable dependiente y es continua en el intervalo unidad, es decir, $0 < y < 1$.

Este capítulo de estructura en 5 secciones. En la primera describimos la estructura del modelo, la inferencia correspondiente se desarrolla en la segunda sección, en la siguiente se presentan los contrastes de hipótesis y las medidas de ajuste en la sección 1.4. Por último, se presenta una posible extensión del modelo clásico.

La bibliografía utilizada para el desarrollo de este capítulo es:

- FERRARI y CRIBARI-NETO (2004). “Beta Regression for Modelling Rates and Proportions”[4]
- CRIBARI-NETO y QUEIROZ (2014). “On testing inference in beta regressions”[5]

Hipótesis distribucional

El modelo se basa en asumir que la variable objetivo y se distribuye según la distribución beta, la cual es apropiada para modelar proporciones, como se ha recogido previamente en el Capítulo 1.

Tomamos una nueva parametrización. Sea $\mu = \frac{\alpha}{\alpha + \beta} y \phi = \alpha + \beta$ se tiene que $\alpha = \mu\phi$ y $\beta = (1 - \mu)\phi$. En consecuencia, la esperanza y la varianza de la variable objetivo se pueden escribir en función de estos nuevos parámetros de la siguiente forma:

$$E[y] = \mu \quad Var[y] = \frac{\mu(1 - \mu)}{1 + \phi}$$

donde μ es la media de la variable respuesta y ϕ puede ser interpretado como el parámetro de precisión, en el sentido de, fijado μ , a mayor valor de ϕ , menor es la varianza de la variable. La función de densidad puede ser escrita con la nueva parametrización como:

$$f(y, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1}(1 - y)^{(1-\mu)\phi-1}, \quad 0 < y < 1$$

donde $0 < \mu < 1$ y $\phi > 0$. El modelo propuesto es también útil para situaciones donde la variable respuesta se restringe a intervalos genéricos (a, b) , donde a y b son escalares $a < b$. En este caso se modela con $\frac{y - a}{b - a}$ en vez de modelar con y directamente.

Hipótesis estructural

Sean y_1, \dots, y_n variables aleatorias independientes, donde para cada y_t , $t = 1, \dots, n$, se asume la hipótesis distribucional anterior con media μ_t y parámetro de precisión ϕ desconocido. El modelo se obtiene asumiendo que la media de cada y_t , μ_t , es vinculado al predictor lineal, η_t , como sigue:

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = x_t^\top \beta = \eta_t$$

donde $\beta = (\beta_1, \dots, \beta_k)^\top$ es un vector desconocido de parámetros de regresión ($\beta \in \mathbb{R}$) y (x_{t1}, \dots, x_{tk}) es un vector de k componentes ($k < n$). Finalmente, $g : (0, 1) \mapsto \mathbb{R}$ es una función vínculo estrictamente creciente y dos veces diferenciable. La varianza de y_t es una función de μ_t y, en consecuencia, de los valores covariantes x_t . Por tanto, la varianza de la variable respuesta puede ser incorporada al modelo.

Algunas posibles elecciones para la función vínculo $g(\cdot)$ son:

- **Función logit:** $g(\mu) = \log \left\{ \frac{\mu}{1 - \mu} \right\}$
- **Función probit:** $g(\mu) = \Phi^{-1}(\mu)$, donde $\Phi(\cdot)$ es la función de distribución de una variable aleatoria normal estándar.
- **Función log-log:** $g(\mu) = -\log \{-\log(1 - \mu)\}$
- **Función log-log complementaria:** $g(\mu) = \log \{-\log(1 - \mu)\}$

En particular, la función vínculo logit se puede escribir:

$$x_t^\top \beta = \log \left\{ \frac{\mu}{1 - \mu} \right\} \iff \mu_t = \frac{e^{x_t^\top \beta}}{1 + e^{x_t^\top \beta}}$$

donde $x_t^\top = (x_{t1}, \dots, x_{tk})$, $t = 1, \dots, n$. Aquí, los parámetros de regresión tienen una interpretación interesante. Supongamos que el valor de la i -ésima variable, x_{ti} , es incrementado en c unidades, manteniendo las demás variables con su valor original, y sea μ^* la media de y bajo los nuevos valores de las variables explicativas y μ la media de y con los valores originales. Entonces, es fácil comprobar que

$$e^{c\beta_i} = \frac{\mu^*/(1 - \mu^*)}{\mu/(1 - \mu)}$$

de manera que $e^{c\beta_i}$ es igual a la *odds ratio*. Para el desarrollo de la siguiente sección, correspondiente a la estimación de parámetros, se ha elegido la función vínculo logit.

2.2. Estimación de parámetros

La estimación de parámetros suele resolverse por el método de máxima verosimilitud. La función de log-verosimilitud basada en una muestra de n variables respuesta independientes distribuidas según una beta es:

$$l(\beta, \phi) = \sum_{t=1}^n l_t(\mu_t, \phi)$$

donde

$$l_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t + ((1 - \mu_t)\phi - 1) \log(1 - y_t)$$

La función score, $U = (U_\beta(\beta, \phi), U_\phi(\beta, \phi))^\top$, es obtenida derivando la función de log-verosimilitud con respecto a los parámetros desconocidos del modelo (β, ϕ) .

$$\frac{\partial l(\beta, \phi)}{\partial \beta_i} = \sum_{t=1}^n \frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i}$$

Notamos que $d\mu_t/d\eta_t = 1/g'(\mu_t)$ y que

$$\frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} = \phi \left[\log \frac{y_t}{1-y_t} - \{\psi(\mu_t\phi) - \psi((1-\mu_t)\phi)\} \right]$$

donde $\psi(\cdot)$ es la función digamma, recogida en el Capítulo 1.

Si llamamos $y_t^* = \log \{y_t/(1-y_t)\}$ y $\mu_t^* = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi)$ se puede escribir la derivada de la función de log-verosimilitud de la siguiente forma:

$$\frac{\partial l(\beta, \phi)}{\partial \beta_i} = \phi \sum_{t=1}^n (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} x_{ti}$$

de la cual obtenemos la expresión matricial de la función score para β :

$$U_\beta(\beta, \phi) = \phi X^\top T(y^* - \mu^*)$$

donde X es una matriz $n \times k$ cuya t -ésima componente es x_t^\top , $y^* = (y_1^*, \dots, y_n^*)^\top$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$ y T es una matriz diagonal cuya expresión es

$$T = \text{diag} \{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$$

Se puede demostrar de manera similar que la función score para ϕ puede ser escrita de la siguiente forma:

$$U_\phi(\beta, \phi) = \sum_{t=1}^n \{\mu_t(y_t^* - \mu_t^*) + \log(1-y_t) - \psi((1-\mu_t)\phi) + \psi(\phi)\}$$

El siguiente paso es obtener la matriz de información de Fisher. La segunda derivada de $l(\beta, \phi)$ con respecto a los parámetros β s viene dada por

$$\begin{aligned} \frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} &= \sum_{t=1}^n \frac{\partial}{\partial \mu_t} \left(\frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_j} x_{ti} \\ &= \sum_{t=1}^n \left(\frac{\partial^2 l_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} + \frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} x_{ti} x_{tj} \end{aligned}$$

Como $E(\partial l_t(\mu_t, \phi)/\partial \mu_t) = 0$, se tiene

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} \right) = \sum_{t=1}^n E \left(\frac{\partial^2 l_t(\mu_t, \phi)}{\partial \mu_t^2} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ti} x_{tj}$$

Ahora, derivando la expresión anterior $\partial l_t(\mu_t, \phi)/\partial \mu_t$ se obtiene

$$\frac{\partial^2 l_t(\mu_t, \phi)}{\partial \mu_t^2} = -\phi^2 \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi)$$

y por tanto, denotando $w_t = \phi \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi) \} \frac{1}{(g'(\mu_t))^2}$,

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} \right) = -\phi \sum_{t=1}^n w_t x_{ti} x_{tj}$$

En forma matricial, se escribe

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta \partial \beta^\top} \right) = -\phi X^\top W X$$

siendo $W = \text{diag} \{w_1, \dots, w_n\}$.

La segunda derivada de $l(\beta, \phi)$ con respecto a algún β_i y ϕ puede ser escrita como

$$\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} = \sum_{t=1}^n \left[(y_t^* - \mu_t^*) - \phi \frac{\partial \mu_t^*}{\partial \phi} \right] \frac{1}{g'(\mu_t)} x_{ti}$$

Si $E(y_t^*) = \mu_t^*$ y $\partial \mu_t^* / \partial \phi = \psi'(\mu_t \phi) \mu_t - \psi'((1 - \mu_t)\phi)(1 - \mu_t)$, se llega a

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = -\sum_{t=1}^n c_t \frac{1}{g'(\mu_t)} x_{ti}$$

donde $c_t = \phi \{ \psi'(\mu_t \phi) \mu_t - \psi'((1 - \mu_t)\phi)(1 - \mu_t) \}$. Sea $c = (c_1, \dots, c_n)^\top$, en notación matricial se escribe

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = -X^\top T c$$

Por último, se obtiene $\partial^2 l(\beta, \phi) / \partial \phi^2$ derivando la expresión de $U_\phi(\beta, \phi)$ respecto de ϕ . Sea $d_t = \psi'(\mu_t \phi) \mu_t^2 + \psi'((1 - \mu_t)\phi)(1 - \mu_t)^2 - \psi'(\phi)$, se llega a

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \phi^2} \right) = -\sum_{t=1}^n d_t$$

que en forma matricial se puede escribir como

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \phi^2} \right) = -\text{tr}(D)$$

siendo $D = \text{diag} \{d_1, \dots, d_n\}$.

Entonces la matriz de información de Fisher viene dada por

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}$$

donde

- $K_{\beta\beta} = \phi X^\top W X$
- $K_{\beta\phi} = K_{\phi\beta}^\top = X^\top T c$
- $K_{\phi\phi} = \text{tr}(D) = \sum_{t=1}^n \psi'(\mu_t \phi) \mu_t^2 + \psi'((1 - \mu_t) \phi) (1 - \mu_t)^2 - \psi'(\phi)$

Se verifica que bajo condiciones de regularidad, cuando el tamaño de muestra es suficientemente grande, se tiene el siguiente comportamiento asintótico de los estimadores de máxima verosimilitud $\hat{\beta}, \hat{\phi}$:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right)$$

siendo K^{-1} la inversa de la matriz de información de Fisher

$$K^{-1}(\beta, \phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}$$

donde

- $K^{\beta\beta} = \frac{1}{\phi} (X^\top W X)^{-1} \left\{ I_k + \frac{X^\top T c c^\top T^\top X (X^\top w X)^{-1}}{\gamma \phi} \right\}$
- $K^{\beta\phi} = (K^{\phi\beta})^\top = -\frac{1}{\gamma \phi} (X^\top W X)^{-1} X^\top T c$
- $K^{\phi\phi} = \gamma^{-1}$

donde I_k es la matriz identidad $k \times k$ y $\gamma = \text{tr}(D) - \phi^{-1} c^\top T^\top X (X^\top W X)^{-1} X^\top T c$.

Los estimadores de máxima verosimilitud de β y ϕ , es decir $\hat{\beta}$ y $\hat{\phi}$, son solución del sistema de ecuaciones

$$\begin{cases} U_\beta(\beta, \phi) = 0 \\ U_\phi(\beta, \phi) = 0 \end{cases}$$

Sin embargo, estas ecuaciones no son lineales, por lo que es necesario resolverlas por maximización numérica de la función de log-verosimilitud usando un algoritmo de optimización no lineal, como métodos Newton o Quasi-Newton; para más detalles, ver Nocedal y Wright (1999)[6]. Para ello, se requiere especificar un valor inicial para las iteraciones, como el estimador mínimo-cuadrático para β obtenido de una regresión lineal de las variables respuesta transformadas $g(y_1), \dots, g(y_n)$ en X , es decir, $(X^\top X)^{-1} X^\top z$, donde $z = (g(y_1), \dots, g(y_n))^\top$. También se necesita un valor inicial para ϕ . A partir de la expresión de la varianza de las variables respuesta transformadas llegamos a la siguiente expresión de ϕ :

$$\text{var}(g(y_t)) = \frac{\mu_t(1 - \mu_t)}{1 + \phi} \Rightarrow \phi = \frac{\mu_t(1 - \mu_t)}{\text{var}(g(y_t))} - 1$$

Observamos que

$$\text{var}(g(y_t)) \approx \text{var} \{g(\mu_t) + (y_t - \mu_t)g'(\mu_t)\} = \text{var}(y_t) \{g'(\mu_t)\}^2$$

y de ahí se tiene que $\text{var}(y_t) \approx \text{var}(g(y_t)) \{g'(\mu_t)\}^{-2}$. Por tanto, un posible valor inicial para ϕ es

$$\frac{1}{n} \sum_{t=1}^n \frac{\check{\mu}_t(1 - \check{\mu}_t)}{\check{\sigma}_t^2} - 1$$

donde

- $\check{\mu}_t = g^{-1}(x_t^\top (X^\top X)^{-1} X^\top z)$
- $\check{\sigma}_t^2 = \frac{\check{e}^\top \check{e}}{(n - k) \{g'(\check{\mu}_t)\}^2}$

con $\check{e} = z - X(X^\top X)^{-1} X^\top z$ el vector de residuos de mínimos cuadrados.

2.3. Contrastes de hipótesis

Consideremos el contraste de hipótesis para el vector de parámetros $\beta = (\beta_1, \dots, \beta_m)^\top$, con $m < k$,

$$\begin{cases} H_0 : \beta = \beta^{(0)} \\ H_1 : \beta \neq \beta^{(0)} \end{cases}$$

donde $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_m^{(0)})^\top$ es un vector de dimensión m dado. El estadístico basado en la log-verosimilitud viene dado por:

$$w_1 = -2l(\tilde{\beta}, \tilde{\phi}) - l(\hat{\beta}, \hat{\phi})$$

donde $l(\beta, \phi)$ es la función de log-verosimilitud y $(\tilde{\beta}^\top, \tilde{\phi})^\top$ es el estimador de máxima verosimilitud de $(\beta^\top, \phi)^\top$ bajo H_0 . Valores grandes de este estadístico significaría una gran diferencia entre la máxima log-verosimilitud con el máximo obtenido asumiendo como cierta la hipótesis nula, con lo que iría a favor del rechazo de H_0 .

Bajo condiciones de regularidad y bajo H_0 se verifica que w_1 sigue asintóticamente una distribución chi-cuadrado con m grados de libertad.

Otros estadísticos usados para el análisis de este contraste son:

El estadístico score o marcador

Este estadístico se basa en el hecho de que la función score se anula en el estimador de máxima verosimilitud, rechazando así la hipótesis nula cuando sea significativamente distinto de cero.

Para describir el estadístico score, denotamos por $U_{1\beta}$ al vector de dimensión m que contiene a los primeros m elementos de la función score para β , cuya expresión matricial dada anteriormente es

$$U_{1\beta}(\beta, \phi) = \phi X^\top T(y^* - \mu^*)$$

y denotamos $K_{11}^{\beta\beta}$ a la matriz $m \times m$ formada por las primeras m filas y columnas de K^{-1} . Se puede demostrar que, realizando una partición de X como $[X_1, X_2]$ siguiendo la partición de β , $U_{1\beta} = \phi X_1^\top T(y^* - \mu^*)$. Por lo que el estadístico Score se puede escribir como:

$$w_2 = \tilde{U}_{1\beta}^\top \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta}$$

donde las tildes indican que las cantidades son evaluadas en el estadístico de máxima verosimilitud restringido a la hipótesis nula. Del mismo modo, en condiciones regulares y bajo H_0 , el estadístico w_2 sigue asintóticamente una distribución chi-cuadrado con m grados de libertad.

El estadístico de Wald

Otra forma de decidir la veracidad o falsedad del contraste es midiendo la distancia ponderada entre el estimador de máxima verosimilitud $\hat{\beta}$ del vector de parámetros β y su vector determinado por la hipótesis nula. Para ello se usa el estadístico de Wald que viene dado por la expresión:

$$w_3 = (\hat{\beta} - \beta^{(0)})^\top (\hat{K}_{11}^{\beta\beta})^{-1} (\hat{\beta} - \beta^{(0)})$$

donde $\hat{K}_{11}^{\beta\beta}$ equivale a $K_{11}^{\beta\beta}$ evaluado en el estimador de máxima verosimilitud. Bajo ciertas condiciones de regularidad y bajo H_0 , este estadístico también sigue asintóticamente una distribución chi-cuadrado de m grados de libertad.

Un caso particular de este contraste es estudiar la significación de los parámetros de regresión β_i , $i = 1, \dots, k$, es decir, hacer el contraste de hipótesis

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

Para ello, se puede usar la raíz cuadrada del estadístico de Wald, es decir, $\hat{\beta}_i/se(\hat{\beta}_i)$, donde $se(\hat{\beta}_i)$ es el error estándar asintótico del estimador de máxima verosimilitud de $\hat{\beta}_i$, obtenido de la inversa de la matriz de información de Fisher evaluada en las estimaciones de máxima verosimilitud. Asintóticamente, este estadístico sigue una distribución normal estándar.

2.4. Medidas de ajuste

Una vez planteado el modelo, es importante analizar la bondad de ajuste, midiendo el grado de explicación de la variabilidad por parte del modelo y la influencia de las variables explicativas en la variable respuesta.

Una medida global de la variación explicada puede ser obtenida por el conocido como "pseudo coeficiente R^2 ", (R_p^2), definido como el cuadrado del coeficiente de correlación muestral entre $\hat{\eta}$ y $g(y)$. Obsérvese que $0 \leq R_p^2 \leq 1$ y $R_p^2 = 1$ indica que hay perfecta concordancia entre $\hat{\eta}$ y $g(y)$ y por tanto entre $\hat{\mu}$ y y .

La desviación del modelo es medida como dos veces la diferencia entre la máxima verosimilitud del modelo saturado y la debida por el modelo bajo estudio. Es decir, el estadístico queda definido como sigue:

$$D(y; \mu, \phi) = \sum_{t=1}^n 2(l_t(\tilde{\mu}_t, \phi) - l_t(\mu_t, \phi))$$

donde $\tilde{\mu}_t$ es el valor de μ_t que resuelve

$$\frac{\partial l_t}{\partial \mu_t} = 0 \Leftrightarrow \phi(y_t^* - \mu_t^*) = 0$$

Cuando ϕ es conocido, esta medida de dispersión es $D(y; \bar{\mu}, \phi)$, donde $\bar{\mu}$ es el estimador de máxima verosimilitud de μ del modelo en particular (con ϕ conocido). Si ϕ es desconocido, el estadístico desviación del modelo es $D(y; \hat{\mu}, \hat{\phi})$, el cual tiene la siguiente expresión:

$$D(y; \hat{\mu}, \hat{\phi}) = \sum_{t=1}^n (r_t^d)^2$$

donde $r_t^d = \text{signo}(y_t - \hat{\mu}_t) 2(l_t(\tilde{\mu}_t, \hat{\phi}) - l_t(\hat{\mu}_t, \hat{\phi}))^{1/2}$ es la contribución de la t-ésima observación a la desviación.

Otras medidas que proporcionan información sobre el ajuste realizado, y el papel de cada una de las observaciones en dicho ajuste, son los residuos estandarizados, definidos como sigue:

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{v\hat{a}r(y_t)}}$$

siendo $\hat{\mu}_t = g^{-1}(x_t^\top \hat{\beta})$ y $v\hat{a}r(y_t) = \hat{\mu}_t(1 - \hat{\mu}_t)/(1 + \hat{\phi})$.

2.5. Otras variantes del modelo

El modelo antes desarrollado mantiene fijo el parámetro de precisión ϕ al variar las observaciones de la variable objetivo. Un modelo desarrollado por Simas permite al parámetro ϕ variar a través de las observaciones, incluyendo así un submodelo para este parámetro:

$$h(\phi_t) = \sum_{j=1}^q z_{tj}\gamma_j = \vartheta_t$$

donde $\gamma = (\gamma_1, \dots, \gamma_q)^\top$ es un vector de parámetros desconocidos ($\gamma \in \mathbb{R}$), z_{t1}, \dots, z_{tq} son q observaciones conocidas y fijas ($q < n - k$), y $h(\cdot) : (0, \infty) \mapsto \mathbb{R}$ es una función vínculo estrictamente creciente y dos-diferenciable. Las función vínculo más usadas para este submodelo son:

$$h(\phi_t) = \log(\phi_t); \quad h(\phi_t) = \sqrt{\phi_t}$$

Capítulo 3

Aplicaciones con R

3.1. El paquete `Betareg`

Todo el modelo teórico descrito en el capítulo anterior es implementado en el programa R mediante un paquete llamado `betareg`, en el cual se encuentran todas las herramientas para realizar la inferencia del modelo clásico de regresión beta, así como para el modelo extendido mencionado en el capítulo anterior. A continuación, se explica cómo usar esta función describiendo los argumentos necesarios.

El contenido de esta sección ha sido obtenido de ZEILEIS Y OTROS (2013). “Package ‘betareg’”[7].

```
betareg(formula, data subset, na.action, weights, offset, link
= c("logit", "probit", "cloglog", "cauchit", "log", "loglog"), link.phi
= NULL, type = c("ML", "BC", "BR"), control = betareg.control(...),
model = TRUE, y = TRUE, x = FALSE, ...)
```

```
betareg.fit(x, y, z = NULL, weights = NULL, offset = NULL, link
= "logit", link.phi = "log", type = "ML", control = betareg.control())
```

Argumentos:

<code>formula</code>	descripción simbólica del modelo (del tipo $y \sim x$ o $y \sim x z$)
<code>data, subset, na.action</code>	argumentos que controlan el procesamiento de la fórmula por <code>model.frame</code> .
<code>weights</code>	vector numérico opcional de ponderaciones de los casos.
<code>offset</code>	vector numérico opcional con una componente conocida para ser incluida en el predictor lineal para la media. En <code>betareg.fit</code> , <code>offset</code> también puede ser una lista de dos <code>offsets</code> para la ecuación de la media y el parámetro de dispersión respectivamente.
<code>link</code>	especifica la función vínculo en el modelo. Actualmente, "logit", "probit", "cloglog", "cauchit", "log", "loglog" están disponibles. Alternativamente, se puede usar un objeto de clase "link-glm".
<code>link.phi</code>	especifica la función vínculo para el modelo del parámetro de dispersión ϕ . Actualmente, "identity", "log", "sqrt" están disponibles. Por defecto se usa "log", a no ser que la fórmula sea del tipo $y \sim x$ donde por defecto es "identity". Alternativamente, se puede usar un objeto de clase "link-glm".
<code>type</code>	especifica el tipo de estimador. Actualmente, están disponibles máxima verosimilitud ("ML"), ML con corrección de sesgo ("BC") y ML con reducción de sesgo ("BR").
<code>control</code> <code>model, y, x</code>	lista de argumentos de control especificados en <code>betareg.control</code> . funciones lógicas. Si aparece <code>TRUE</code> se devuelven las componentes correspondientes al ajuste del modelo (estructura del modelo función de la variable respuesta, matriz del modelo). Para <code>betareg.fit</code> , <code>x</code> es la matriz numérica correspondiente a los valores covariantes e <code>y</code> es el vector respuesta con valores en (0,1).
<code>z</code>	matriz numérica. Matriz de regresión para el modelo de precisión (modelo para el parámetro ϕ).
<code>...</code>	argumentos de <code>betareg.control</code> .

La función `betareg` emplea el comando `optim` para la estimación por el método de máxima verosimilitud, el cual usa gradientes analíticos y valores iniciales de una regresión lineal auxiliar de la variable respuesta transformada.

`betareg.fit` devuelve una lista con componentes de `converged`. Finalmente, `betareg` devuelve una lista con las componentes siguientes:

<code>coefficients</code>	una lista de elementos "mean" y "precision" que contienen los coeficientes del modelo para la media y el parámetro de precisión respectivamente.
<code>residuals</code>	un vector de residuos brutos, es decir, diferencia entre los residuos observados y los obtenidos por el modelo.
<code>fitted.values</code>	vector de medias estimado.
<code>optim</code>	resultado de <code>optim</code> , comando usado para la estimación por máxima log-verosimilitud.
<code>method</code>	argumento <code>method</code> del comando <code>optim</code>
<code>control</code>	argumento <code>control</code> del comando <code>optim</code>
<code>start</code>	valores iniciales de los parámetros usados por <code>optim</code> .
<code>weights</code>	ponderaciones usadas (si las hay).
<code>offset</code>	lista de vectores <code>offset</code> usados (si los hay)
<code>n</code>	número de observaciones.
<code>nobs</code>	número de observaciones con ponderaciones no nulas.
<code>df.null</code>	grados de libertad en el modelo nulo (media y dispersión constantes), es decir, $n-2$.
<code>df.residual</code>	grados de libertad en el modelo ajustado.
<code>phi</code>	función lógica que indica si los coeficientes del parámetro de precisión serán tratados como parámetros del modelo.
<code>loglik</code>	log-verosimilitud del modelo.
<code>vcov</code>	matriz de covarianzas de todos los parámetros del modelo.
<code>pseudo.r.squared</code>	valor del cuadrado del coeficiente de correlación del predictor lineal y la variable respuesta transformada (pseudo R^2).
<code>link</code>	lista de elementos "mean" "precision" que contienen las especificaciones de los vínculos en los respectivos modelos.

<code>converged</code>	función lógica que indica si <code>optim</code> converge.
<code>call</code>	función original <code>call</code> .
<code>formula</code>	fórmula original.
<code>terms</code>	lista con los elementos "mean", "precisionz" "full" que contienen las condiciones de los respectivos modelos.
<code>levels</code>	lista con los elementos "mean", "precision" y "full" que contienen los niveles de las variables categóricas.
<code>contrasts</code>	lista con los elementos "mean", "precision" que contienen los contrastes correspondientes a los niveles de los respectivos modelos.
<code>model</code>	estructura del modelo en caso de <code>model=TRUE</code> .
<code>y</code>	vector de la variable respuesta en caso de <code>y=TRUE</code>
<code>x</code>	lista con los elementos "mean", "precision" que contienen las matrices de los respectivos modelos, en caso de <code>x=TRUE</code>

Comando `beta.control`

Todos los parámetros en `betareg` son estimados por máxima verosimilitud usando `optim` con opciones de control en `beta.control`. Los valores iniciales pueden ser elegidos via `start` o estimados por `lm.wfit`.

```
betareg.control(phi = TRUE, method = "BFGS", maxit = 5000, hessian
= FALSE, trace = FALSE, start = NULL, fsmaxit = 200, fstol = 1e-8,
...)
```

Argumentos:

<code>phi</code>	indica si el parámetro de precisión ϕ debe ser tratado como un parámetro del modelo (<code>TRUE</code> por defecto).
<code>method</code>	lista de caracteres que especifica el argumento <code>method</code> de <code>optim</code>

<code>maxit</code>	número entero que especifica el argumento <code>maxit</code> (máximo número de iteraciones) de <code>optim</code> .
<code>trace</code>	argumento lógico o número entero que controla si la información durante el progreso de la optimización debe ser mostrada.
<code>hessian</code>	argumento lógico que controla si la matriz hessiana resultante de <code>optim</code> debe ser usada para la estimación de la matriz de covarianzas. Por defecto, es empleada.
<code>start</code>	vector opcional con valores iniciales para todos los parámetros, incluido el parámetro ϕ .
<code>fsmxit</code>	número máximo de iteraciones en el método iterativo marcador de Fisher.
<code>fstol</code>	tolerancia numérica para la convergencia del método iterativo de Fisher.
<code>...</code>	argumentos para <code>optim</code> .

Plot en `betareg`

El comando `plot` para objetos `betareg` produce varios tipos de gráficos de diagnóstico, incluyendo varios tipos de residuos, medidas de influencia, etc.

```
plot(x, which = 1:4, caption = c("Residuals vs indices of obs.",
"Cook's distance plot", "Generalized leverage vs predicted values",
"Residuals vs linear predictor", "Half-normal plot of residuals",
"Predicted vs observed values"), sub.caption = paste(deparse(x$call),
collapse = " @"), main = @, ask = prod(par("mfcoll")) <length(which)
&& dev.interactive(), ..., type = "sweighted2", nsim = 100, level
= 0.9)
```

Argumentos:

<code>x</code>	modelo objeto de clase " <code>betareg</code> "
<code>which</code>	argumento numérico. Si sólo es necesario un subconjunto de gráficos, especifica un subconjunto de los números 1:6.
<code>caption</code>	caracteres. Comentario al pie de la gráfica.
<code>sub.caption</code>	caracteres. Título común bajo las figuras en caso de haber múltiples.
<code>main</code>	caracteres. Título de cada gráfica además del pie de imagen.
<code>ask</code>	argumento lógico. Si <code>TRUE</code> , se pregunta al usuario antes de cada gráfica.
<code>...</code>	otros parámetros para representar funciones.
<code>type</code>	carácter que indica el tipo de residuo que se usa, ver <code>residuals.betareg</code> .
<code>nsim</code>	numérico. Número de simulaciones en half-normal plots.
<code>level</code>	numérico. Nivel de confianza en half-normal plots.

Predicciones en `betareg`

Las predicciones en un modelo de regresión beta pueden darse en valores del intervalo (0,1) o en una escala del predictor lineal. En R, usamos la función `predict`.

```
predict(object, newdata = NULL, type = c("response", "link", "precision",  
"variance", "quantile"), na.action = na.pass, at = 0.5, ...)
```

Argumentos:

<code>object</code>	modelo objeto de clase " <code>betareg</code> "
<code>newdata</code>	opcionalmente, es una base de datos para buscar variables para predecir. Si se omite, se usa las observaciones originales.
<code>type</code>	carácter que indica el tipo de predicciones: correspondiente a la media de las variables respuesta (" <code>response</code> "), al predictor lineal (" <code>link</code> "), al parámetro de precisión (" <code>precision</code> "), a la varianza de las variables respuesta (" <code>variance</code> "), o a los cuantiles de la distribución de las variables respuesta (" <code>quantile</code> ")
<code>na.action</code>	función que determina lo que hacer con los valores perdidos en <code>newdata</code> .
<code>at</code>	vector numérico que indica los cuantiles para predecir en caso de <code>type = "quantile"</code> . Por defecto, <code>at = 0.5</code> .

Base de datos `GasolineYield`.

A continuación, se usa el paquete `betareg` para aplicar el modelo de regresión beta a una base de datos llamada `GasolineYield`, que contiene las proporciones de petróleo convertido en gasolina tras la destilación fraccionada. Este conjunto de datos esta formado por 32 observaciones de 6 variables:

<code>yield</code>	proporción de petróleo convertido el gasolina tras la destilación fraccionada.
<code>gravity</code>	gravedad del petróleo (en grados API).
<code>pressure</code>	presión del petróleo (lbf/in ²).
<code>temp10</code>	temperatura (en grados F) en la cual el 10 % del petróleo se ha vaporizado.
<code>temp</code>	temperatura (en grados F) en la cual todo el petróleo se ha vaporizado.
<code>batch</code>	factor que indica el conjunto de condiciones <code>gravity</code> , <code>pressure</code> y <code>temp10</code> .

La variable dependiente es la proporción de petróleo convertido en gasolina, siendo el resto las variables explicativas del modelo. Diez subconjuntos de valores de las tres primeras variables explicativas corresponden a diez tipos de petróleo distintos, los cuales fueron sujetos en la destilación a condiciones controladas experimentalmente. Estas condiciones se recogen en la variable batch.

Código:

```
>data("GasolineYield", package = "betareg")
>gy <- betareg(yield gravity + pressure + temp10 + temp, data =
GasolineYield)
>summary(gy)
```

Los resultados del código anterior son los siguientes:

Call:

```
betareg(formula = yield ~ gravity + pressure + temp10 + temp, data =
GasolineYield)
```

Standardized weighted residuals 2:

Min	1Q	Median	3Q	Max
-2.1189	-0.6985	-0.0088	0.6306	2.1572

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.6949422	0.7625693	-3.534	0.000409	***
gravity	0.0045412	0.0071419	0.636	0.524871	
pressure	0.0304135	0.0281007	1.082	0.279117	
temp10	-0.0110449	0.0022640	-4.879	1.07e-06	***
temp	0.0105650	0.0005154	20.499	<2e-16	***

Phi coefficients (precision model with identity link):

	Estimate	Std. Error	z value	Pr(> z)	
(phi)	248.24	62.02	4.003	6.26e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)

Log-likelihood: 75.68 on 6 Df

Pseudo R-squared: 0.9398

Number of iterations: 144 (BFGS) + 5 (Fisher scoring)

Por tanto, los resultados del modelo indica la significación de las variables

`temp10` y `temp` en la modelización de la variable `yield`. Y sin embargo, la no significación de las variables `gravity` y `pressure`.

La representación gráfica de los resultados se obtiene a través de las siguientes órdenes:

```
>par(mfrow = c(3, 2))  
>plot(gy, which = 1:6)  
>par(mfrow = c(1, 1))
```

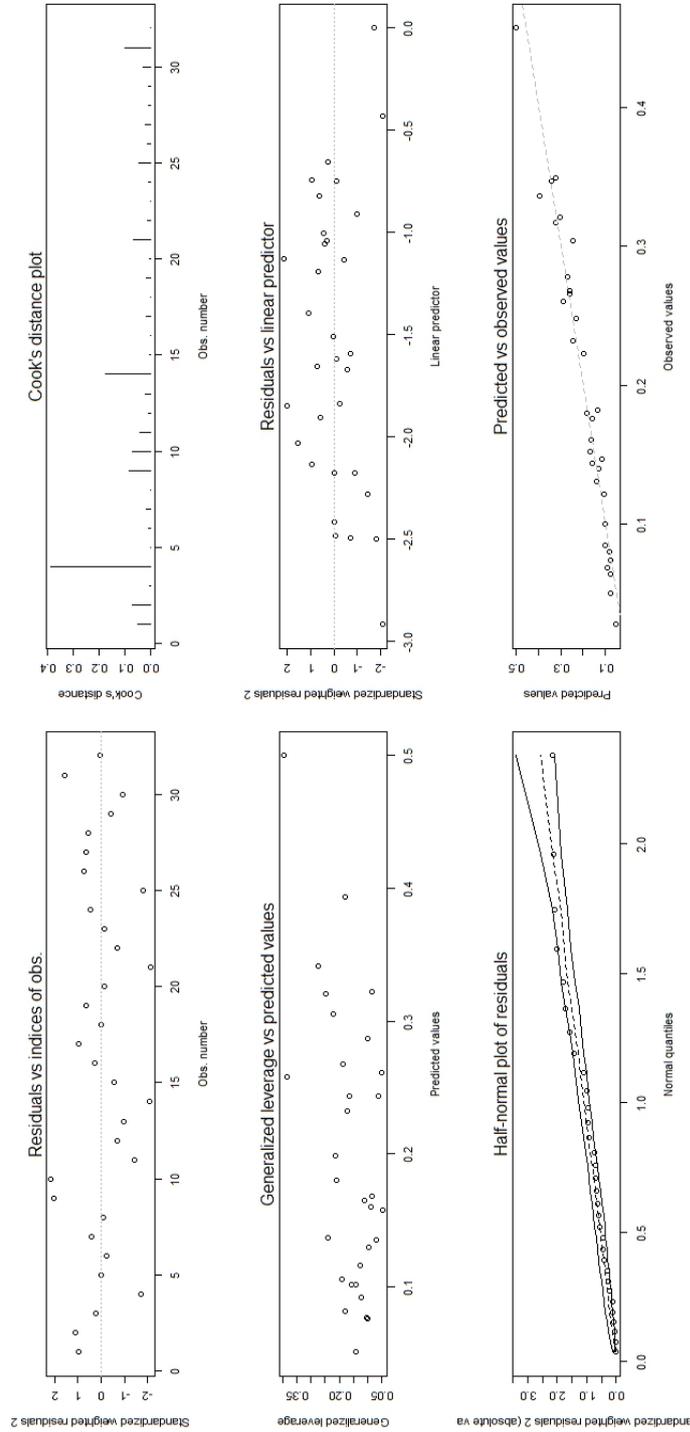


Figura 3.1: Conjunto de gráficas generadas con el comando plot

3.2. Aplicaciones prácticas

En esta sección se mencionan y describen algunos estudios reales relacionados con el modelo de regresión beta:

- BRETON Y OTROS (2014). “Prenatal Tobacco Smoke Exposure Is Associated with Childhood DNA CpG Methylation”[8]

Fumar durante el embarazo puede provocar múltiples consecuencias negativas en la salud del niño. El objetivo de este artículo es relacionar modificaciones epigenéticas en niños asmáticos con exposición prenatal al tabaco. Para ello, usa el modelo de regresión beta con variable objetivo el porcentaje de metilación del ADN en una muestra de sangre y como variables predictoras medida de exposición al tabaco en el útero, edad, sexo, clínica de atención prenatal y tipo de célula estimada. Los autores de esta investigación concluyen que fumar durante el embarazo está relacionado con cambios epigenéticos que persisten durante la infancia.

- SWEARINGEN Y OTROS (2011). “Application of Beta Regression to Analyze Ischemic Stroke Volume in NINDS rt-PA Clinical Trials”[9]

La cantidad de lesión ictus isquémica es difícil de analizar debido a la forma de su distribución subyacente. El estudio consiste en ver si la regresión beta es un buen modelo para estudiar esta variable. Para ello comparan los resultados obtenidos en este modelo con un análisis publicado con anterioridad. Finalmente, todas las variables previamente identificadas de importancia fueron detectadas en el modelo construido concluyendo así que el modelo de regresión beta es una alternativa legítima para analizar la cantidad de lesión provocada por un ictus isquémico.

- HUNGER Y OTROS (2011). “Analysis of SF-6D Index Data: Is Beta Regression Appropriate?”[10]

El pino de corteza blanca (*Pinus albicaulis*) depende del cascanueces americano (*Nucifraga columbiana*) para la dispersión de sus semillas. Mientras el pino de corteza blanca sucumbe a una enfermedad exótica y a los escarabajos del pino de montaña (*Dendroctonus ponderosae*), disminuye la producción de piña, por lo que los cascanueces americanos los visitan con menos frecuencia, disminuyendo la probabilidad de dispersar sus semillas. Este estudio de investigación demuestra que

el modelo de regresión beta es el método estadístico más apropiado para modelar la relación entre la densidad de piña y la proporción del tiempo de estudio en el que se observan cascanueces americanos.

Bibliografía

- [1] KIESCHNICK y MCCULLOUGH (2003). “Regression analysis of variates observed on (0, 1): percentages, proportions and fractions”, *Statistical Modelling* 3, 193–213.
- [2] JOHNSON Y OTROS (1995). “Chapter 21: Beta Distributions”, *Continuous Univariate Distributions Vol. 2* (2nd ed.). Wiley. ISBN 978-0-471-58494-0.
- [3] WIKIPEDIA CONTRIBUTORS (2015). “Beta distribution”, *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 28 Aug. 2015. Web. 5 Sep. 2015.
- [4] FERRARI y CRIBARI-NETO (2004). “Beta Regression for Modelling Rates and Proportions”, *Journal of Applied Statistics* 31:7, 799-815, DOI: 10.1080/0266476042000214501.
- [5] CRIBARI-NETO y QUEIROZ (2014). “On testing inference in beta regressions”, *Journal of Statistical Computation and Simulation* 84:1, 186-203, DOI: 10.1080/00949655.2012.700456.
- [6] NOCEDAL y WRIGHT (1999). “Numerical Optimization”, *New York: Springer-Verlag*
- [7] ZEILEIS Y OTROS (2013). “Package ‘betareg’”, Beta Regression, *CRAN*, 2013-09-16 18:59:12
- [8] BRETON Y OTROS (2014). “Prenatal Tobacco Smoke Exposure Is Associated with Childhood DNA CpG Methylation”, *PLoS ONE* 9(6): e99716.
- [9] SWEARINGEN Y OTROS (2011). “Application of Beta Regression to Analyze Ischemic Stroke Volume in NINDS rt-PA Clinical Trials”, *Neuroepidemiology*, 37:73–82.
- [10] HUNGER Y OTROS (2011). “Analysis of SF-6D Index Data: Is Beta Regression Appropriate?”, *Value in Health*, 759–767