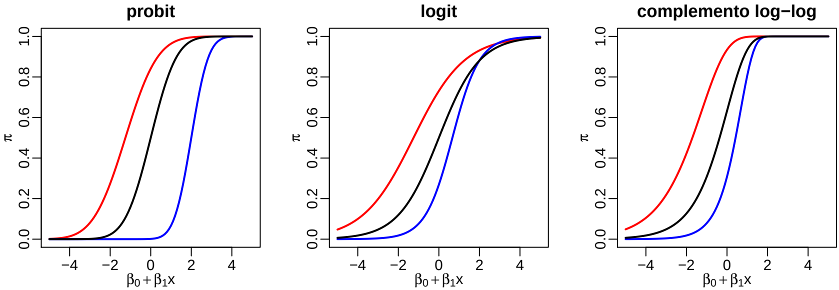
MLG – Regresión Logística



−ebz

Contenido

Contenido

1. Modelos lineales generalizados (MLG).
   - Ejemplos.
   - Familia exponencial.
   - Componentes de un modelo MLG.
   - Estimación: máxima verosimilitud; mínimos cuadrados.
   - Bondad de ajuste: devianza.
   - Inferencia.
   - Residuales.
2. Regresión logística.
   - Respuesta binaria.
   - Modelos dosis-respuesta.
   - Regresión logística.
   - Interpretación.
   - Inferencia.
   - Ejemplo.

a) Respuesta binaria

Ejemplo: Problema de recién nacidos bajos de peso[1]

El problema de niños con bajo peso al nacer es una condición que preocupa a los médicos pues constituye un grupo de alto riesgo para los problemas de mortalidad infantil y tasas de defectos al nacer. El comportamiento de la madre durante el embarazo (dieta, hábitos de fumar, cuidado prenatal, etc.) puede afectar de manera importante el tiempo de gestación y en consecuencia el peso normal de un niño al nacer.

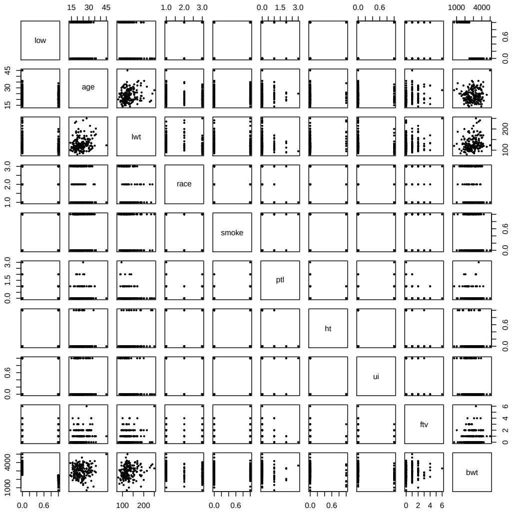
variable	código	etiqueta
Código de identificación:	...	id
Bajo peso:	0 = peso$\geq$2500 gr. 1 = peso$<$2500 gr.	low
Edad de la madre:	años	age
Peso antes de la última menstruación:	libras	lwt
Raza:	1=blanca;2=negra;3=otra	race
Hábito del fumar durante embarazo:	1=sí;0=no	smoke
Historia de labor prematura	0=ninguna;1=una;etc.	ptl
Historia de hipertensión	1=sí;0=no	ht
Irritación uterina:	1=sí;0=no	ui
Número de visitas médicas durante el primer trimestre:	0=ninguna;1=una;etc.	ftv
Peso al nacer:	gramos	bwt

obs	id	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt
1	4	1	28	120	3	1	1	0	1	0	709
2	10	1	29	130	1	0	0	0	1	2	1021
3	11	1	34	187	2	1	0	1	0	0	1135
4	13	1	25	105	3	0	1	1	0	0	1330
5	15	1	25	85	3	0	0	0	1	0	1474
6	16	1	27	150	3	0	0	0	0	0	1588
.	.	.	.	.	.	.	.	.	.	.	.

[1] Hosmer and Lemeshow (1989)

−ebz

Problema de recién nacidos bajos de peso (cont.)

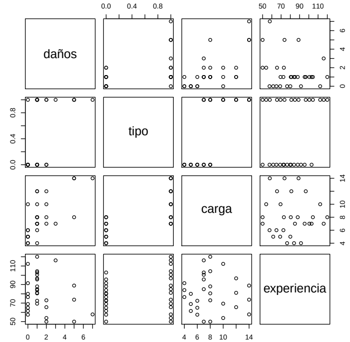


b) Respuesta conteo

***Ejemplo*: Número de daños a aviones**[2]

Durante la guerra de Vietnam la marina de los E. U. usó varios tipos de aviones de ataque, en ocasiones en acciones de baja altura contra construcciones y transporte. Dos de los aviones fueron los A4 construcción de McDonnell Douglas y los A6 de Grumman. Se tiene una muestra de 30 misiones de ataque donde intervinieron ambos tipos de bombarderos. La variable respuesta de interés y es la cantidad de lugares donde se infligieron daños al avión. Las variables explicativas son: x_1 indicadora del tipo de avión (A4=0, A6=1); x_2 la carga en toneladas del avión; y x_3 la experiencia en meses de la tripulación.

obs	y	x1	x2	x3	obs	y	x1	x2	x3
1	0	0	4	91.5	16	3	1	7	116.1
2	1	0	4	84.0	17	1	1	7	100.6
3	0	0	4	76.5	18	1	1	7	85.0
4	0	0	5	69.0	19	1	1	10	69.4
5	0	0	5	61.5	20	2	1	10	53.9
6	0	0	5	80.0	21	0	1	10	112.3
7	1	0	6	72.5	22	1	1	12	96.7
8	0	0	6	65.0	23	1	1	12	81.1
9	0	0	6	57.5	24	2	1	12	65.6
10	2	0	7	50.0	25	5	1	8	50.0
11	1	0	7	103.0	26	1	1	8	120.0
12	1	0	7	95.5	27	1	1	8	104.4
13	1	0	8	88.0	28	5	1	14	88.9
14	1	0	8	80.5	29	5	1	14	73.7
15	2	0	8	73.0	30	7	1	14	57.8



−ebz

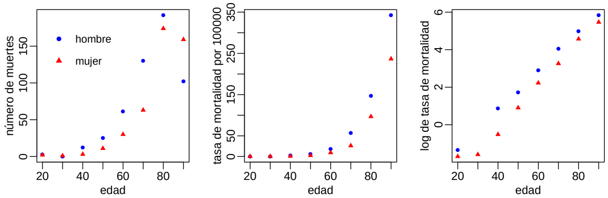
[2] Montgomery et al. (2002)

b) Respuesta conteo

Ejemplo: Muertes por diabetes[3]

En este ejemplo se estudia el número de muertes debida a la diabetes en una región de Australia en 2002. Como variables independientes se consideran el sexo x_1 y la edad x_2.

gender	age	deaths	popn	l_popn	agemidpt	gender	age	deaths	popn	l_popn	agemidpt
Male	<25	3	1141100	13.94750	20	Female	<25	2	1086408	13.89839	20
Male	25-34	0	485571	13.09308	30	Female	25-34	1	489948	13.10205	30
Male	35-44	12	504312	13.13095	40	Female	35-44	3	504030	13.13039	40
Male	45-54	25	447315	13.01102	50	Female	45-54	11	445763	13.00754	50
Male	55-64	61	330902	12.70958	60	Female	55-64	30	323669	12.68748	60
Male	65-74	130	226403	12.33007	70	Female	65-74	63	241488	12.39458	70
Male	75-84	192	130527	11.77934	80	Female	75-84	174	179686	12.09897	80
Male	85+	102	29785	10.30176	90	Female	85+	159	67203	11.11547	90



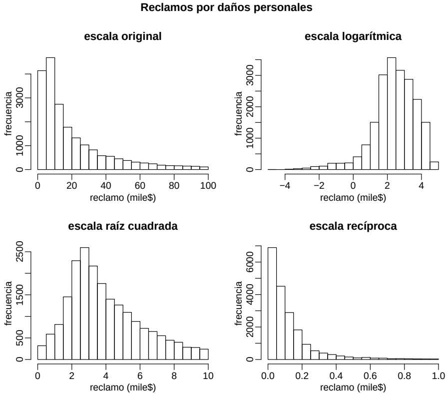
[3] Jong and Heller (2008)

c) Respuesta continua

Ejemplo: Problema reclamos por daños personales[4]

El problema considera 22036 acuerdos por seguro por daños personales en accidentes. La respuesta, total acordado, sigue aproximadamente una distribución *gamma*, que no admite transformaciones básicas.

```
Variables:

total         settled amount
inj1, ...,    injury 1,..., injury 5 coded as
              1 = no injury
              2, 3, 4, 5 = injury severities
              6 = fatal injury
                  9 = not recorded
legrep        legal representation
              (0 = no, 1 = yes)
accmonth      accident month
              (1=July 1989,..., 120=June 1999)
repmonth      reporting month (as above)
finmonth      finalization month (as above)
op_time       operational time
```



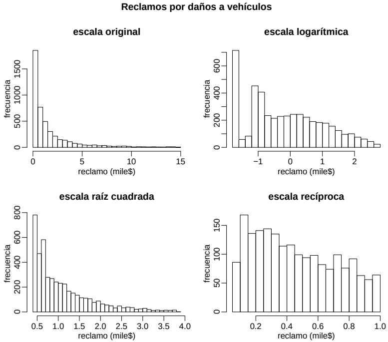
[4] Jong and Heller (2008)

c) Respuesta continua

Ejemplo: Problema reclamos por daños a vehículos[5]

El problema considera 4624 pólizas de seguro por daños a vehículos en accidentes con al menos una reclamación. La respuesta, total pagado, sigue aproximadamente una distribución *inversa gaussiana*, que no admite transformaciones básicas.

```
Variables:

veh_value    vehicle value, in $10,000s
exposure     0-1
clm          occurrence of claim
             (0 = no, 1 = yes)
numclaims    number of claims
claimcst0    claim amount (0 if no claim)
veh_body     vehicle body, coded as
               BUS
               CONVT = convertible
               COUPE
               HBACK = hatchback
               MIBUS = minibus
               PANVN = panel van
               STNWG = station wagon
               TRUCK
               etc.
veh_age      age of vehicle:
                 1 (youngest), 2, 3, 4
gender       gender of driver: M, F
area         driver's area of residence:
                 A, B, C, D, E, F
agecat       driver's age category:
                 1 (youngest), 2, 3, 4, 5, 6
```



[5] Jong and Heller (2008)

Familia Exponencial de Distribuciones

La variable aleatoria Y se dice que es miembro de la *familia exponencial de distribuciones* si su función de densidad de probabilidad, $f(y;\theta)$, puede expresarse como

$$f(y;\theta) = \exp\left\{a(y)b(\theta) + c(\theta) + d(y)\right\} \tag{1}$$

si $a(y) = y$, la distribución anterior (1) se dice estar en su *forma canónica* y a $b(\theta)$ se le llama el *parámetro natural* de la distribución.
Ejemplos:

$$\text{Binomial}: \quad f(y;n,p) = \begin{cases} \binom{n}{y}\pi^y(1-\pi)^{n-y} \\ \exp\left\{y\log\pi - y\log(1-\pi) + n\log(1-\pi) + \log\binom{n}{y}\right\} \end{cases}$$

$$\text{Poisson}: \quad f(y;\lambda) = \begin{cases} \frac{\lambda^y e^{-\lambda}}{y!} \\ \exp\left\{y\log\lambda - \lambda - \log\lambda\right\} \end{cases}$$

$$\text{Normal:}: \quad f(y;\mu,\sigma^2) = \begin{cases} \frac{1}{2\pi\sigma^2}\, e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \\ \exp\left\{-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \end{cases}$$

También incluye otras distribuciones como la *gamma*, la *lognormal*, la *gaussiana inversa*, etc.

−ebz

Reparametrización

Sea *y v. a.* con *f. d. p.* $f(y, \boldsymbol{\theta})$ miembro de la familia exponencial. Entonces como alternativa a la expresión (1), f se puede escribir como

$$f(y;\boldsymbol{\theta}) = \exp\left\{[y\theta - b(\theta)]\frac{1}{a(\phi)} + c(y;\phi)\right\} \tag{2}$$

donde θ es el parámetro *natural o canónico* de *localización* y ϕ el parámetro de *dispersión*.

Distribución	Soporte	θ	$a(\cdot)$	$b(\cdot)$	$c(\cdot)$	$\mu = \mathbb{E}[Y]$
Binomial	$[\![1, n]\!]/n$	$\log\left(\frac{p}{1-p}\right)$	$1/n$	$\log(1+e^{\theta})$	$\log\left[\binom{n}{ny}\right]$	$e^{\theta}/(1+e^{\theta})$
Poisson	$[\![0, \infty]\!]$	$\log(\lambda)$	1	e^{θ}	$-\log y!$	e^{θ}
Binomial Negativa	$[\![0, \infty]\!]$	$\log(1-p)$	1	$-r\log(1-e^{\theta})$	$\log\left[\binom{r+y-1}{y}\right]$	$re^{\theta}/(1-e^{\theta})$
Normal	$(-\infty, \infty)$	μ	ϕ	$\theta^2/2$	$-\frac{1}{2}\left(y^2/\phi - \log(2\pi\phi)\right)$	θ
Gamma	$(0, \infty)$	$-\beta$	ϕ	$\log(-\theta)$	$(\phi^{-1}-1)\left[\log(y\phi)+\log(\phi)\right] - \log\Gamma(\phi^{-1})$	$1/\theta$
Gaussiana Inversa	$(0, \infty)$	$-1/2\mu^2$	ϕ	$-(-2\theta)^{1/2}$	$-\frac{1}{2}\left[1/y\phi - \log(-2\pi\phi y^3)\right]$	$(-2\theta)^{-1/2}$

−ebz

Función de verosimilitud, *score*, etc.

Se define la *función de verosimilitud* por

$$L(\theta; y) = f(y; \theta) = \exp\left\{a(y)b(\theta) + c(\theta) + d(y)\right\}$$

Y la *función log-de-verosimilitud* por

$$\ell(\theta; y) = \log L(\theta; y) = a(y)b(\theta) + c(\theta) + d(y)$$

Se define la *función score* por

$$s(\theta; y) = \frac{\partial \ell(\theta; y)}{\partial \theta} = a(y)b'(\theta) + c'(\theta)$$

entonces se puede mostrar que

$$\begin{aligned} \mathbb{E}[s(\theta; y)] &= \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)\right] \\ &= b'(\theta)\mathbb{E}[a(y)] + c'(\theta) \\ &= 0 \\ \operatorname{var}(s(\theta; y)) &= \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = -\mathbb{E}\left[\left(\frac{\partial^2 \ell}{\partial \theta^2}\right)\right] \\ &= \left[b'(\theta)\right]^2 \operatorname{var}(a(y)) \\ &= I(\theta; y) \end{aligned}$$

–ebz

Componentes de un modelo lineal generalizado

1. Se supone que las *variables respuesta*, $y_1, \ldots, y_n$, siguen una distribución común miembro de la *familia exponencial*.
2. Un conjunto de *variables explicativas*, $x_1, \ldots, x_p$, y de parámetros $\beta_0, \beta_1, \ldots, \beta_p$. Así,

$$y_{n\times 1} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}; \quad \beta_{q\times 1} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}; \quad X_{n\times q} = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

3. Una *función liga* monótona g tal que

$$g(\mu_i) = x_i'\beta$$

donde $\mu_i = \mathbb{E}[y_i]$.

Estimación por Máxima verosimilitud

Dada la muestra $y_1, \ldots, y_n$ de $y \sim f(y;\boldsymbol{\theta})$, con $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_m)$, el estimador $\hat{\boldsymbol{\theta}}$ obtenido por el *método de máxima verosimilitud* es aquel tal que

$$L(\hat{\boldsymbol{\theta}}; y) \geq L(\boldsymbol{\theta}; y), \qquad \text{para todo } \boldsymbol{\theta} \in \Theta$$

Equivalentemente, puesto *log* es una función monótona creciente

$$\ell(\hat{\boldsymbol{\theta}}; y) \geq \ell(\boldsymbol{\theta}; y), \qquad \text{para todo } \boldsymbol{\theta} \in \Theta$$

Generalmente el EMV $\hat{\boldsymbol{\theta}}$ se obtiene por diferenciación de la función log-de-verosimilitud $\ell(\theta; y)$ y resolviendo el sistema

$$\frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial \theta_j} = 0, \qquad \text{para todo } j = 1, \ldots, m$$

Es necesario confirmar que la solución $\hat{\boldsymbol{\theta}}$ corresponde a un máximo verificando que la matriz de segundas derivadas

$$\left. \frac{\partial^2 \ell(\boldsymbol{\theta}; y)}{\partial \theta_j \partial \theta_k} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

es definida negativa.
$\star$ Propiedades de los estimadores de máxima verosimilitud: invarianza, consistencia, suficiencia, eficiencia asintótica.

−ebz

Estimación – Mínimos cuadrados

Suponga que la respuesta y_i es tal que $\mu_i = \mathbb{E}[y_i]$ y μ_i función de los parámetros β. Esto es, $\mu_i = \mu_i(\beta)$. De esta manera, considere

$$y_i = \mu_i + \epsilon_i, \qquad i = 1, \ldots, n$$

La *suma de cuadrados* está dada por

$$S(\beta) = \sum_{i=1}^{n} (y_i - \mu_i(\beta))^2 = (y - \mu)'(y - \mu)$$

Generalmente los EMC de β se obtienen por diferenciación y resolviendo el sistema

$$\frac{\partial S(\beta)}{\partial \beta_j} = 0, \qquad j = 0, \ldots, p$$

Si por alguna razón se desea *ponderar* los términos, por ejemplo por diferencia de varianzas, los estimadores se obtienen minimizado la suma

$$S_W(\beta) = \sum_{i=1}^{n} w_i (y_i - \mu_i(\beta))^2$$

donde $w_i = 1/\text{var}(y_i)$. En el caso que $\mu = X\beta$, la suma de cuadrados es $S_W(\beta) = (y - X\beta)' V^{-1} (y - X\beta)$, que da lugar a las ecuaciones normales

$$X' V^{-1} X \beta = X' V^{-1} y$$

que corresponde a los *mínimos cuadrados ponderados* (MCP).

Modelos lineales generalizados

Considere la muestra $y_1, \ldots, y_n$ de $y \sim f(y; \boldsymbol{\theta})$, miembro de la familia exponencial. Entonces,

$$\ell(\theta; y) = \sum y_i b(\theta_j) + \sum c(\theta_j) + \sum d(y_i)$$

En el caso de *modelos lineales generalizados*,

$$\mu_i = \mathbb{E}[y_i] = -c'(\theta)/b'(\theta)$$
$$\eta_i = g(\mu_i) = x_i' \beta$$

En el caso de la familia exponencial $\hat{\theta}$ puede obtenerse por diferenciación de ℓ, o equivalentemente, resolviendo las ecuaciones $\frac{\partial \ell}{\partial \beta} = 0$.
Se puede mostrar que tiene

$$\frac{\partial \ell}{\partial \beta_j} = s_j = \sum \frac{(y_i - \mu_i) x_{ij}}{\mathrm{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

Ahora bien las ecuaciones, $s_j = 0$, son en general no lineales por lo que hay necesidad de resolverlas numéricamente. Si se emplea el *método de Newton-Raphson*, la r-ésima aproximación está dada por

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} - \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right]^{-1}_{\beta = \hat{\beta}^{(r-1)}} s^{(r-1)} \tag{3}$$

donde $s_j^{(r-1)} = \partial \ell / \partial \beta_j$ evaluado en $\beta = \hat{\beta}^{(r-1)}$.

Mínimos cuadrados re-ponderados iterativos

En ocasiones el *método de scoring* es más simple que *Newton-Raphson*. En la ecuación (3) se sustituye la matriz de segundas derivadas por el negativo de la *matriz de información* $\mathcal{I}$

$$-\mathcal{I}_{jk} = \mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta_j \beta_k}\right] = \mathbb{E}\left[\frac{\partial \ell}{\partial \beta_j} \cdot \frac{\partial \ell}{\partial \beta_k}\right]$$

Luego, la correspondiente expresión para $\hat{\beta}$

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} + \left[\mathcal{I}^{(r-1)}\right]^{-1} s^{(r-1)} \tag{4}$$

o bien,

$$\mathcal{I}^{(r-1)} \hat{\beta}^{(r)} = \mathcal{I}^{(r-1)} \hat{\beta}^{(r-1)} + s^{(r-1)} \tag{5}$$

En el caso de MLG

$$\mathcal{I}_{jk} = \sum_{i=1}^{n} \frac{x_{ij} x_{ik}}{\mathrm{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

por lo que

$$\mathcal{I} = X' W X$$

donde $W = \mathrm{diag}\{w_1, \ldots, w_n\}$ con

$$w_{ii} = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 / \mathrm{var}(y_i)$$

–ebz

Mínimos cuadrados re-ponderados iterativos

Si se define

$$z_i = \sum_j x_{ij}\hat{\beta}^{(r-1)} + (y_i - \mu_i)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$$

donde μ_i y $\partial\eta_i/\partial\mu_i$ son evaluadas en $\hat{\beta}^{(r-1)}$. Luego la ecuación (3) puede escribirse como

$$X'WX\hat{\beta}^{(r)} = X'Wz \tag{6}$$

que tiene la misma forma de las ecuaciones normales para los mínimos cuadrados ponderados. La solución de la ecuación (4)

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} + \left[\mathcal{I}^{(r-1)}\right]^{-1} s^{(r-1)}$$

de manera iterativa y define al método de *mínimos cuadrados re-ponderados iterativos*, pues en general z y W dependen de $\hat{\beta}$.

–ebz

Inferencia – Resultados asintóticos de EMV $\hat{\theta}$

La idea básica es que el EMV $\hat{\theta}$ es un estimador consistente del parámetro θ y que si var$(\hat{\theta})$ es su varianza, entonces:

1. El estimador $\hat{\theta}$ es *asintóticamente insesgado*.

$$\mathbb{E}[\hat{\theta}_n] \to \theta$$

2. El estadístico $\hat{\theta}$ tiene distribución asintótica normal.

$$\frac{\hat{\theta}-\theta}{\sqrt{\mathrm{var}(\hat{\theta})}} \mathrel{\dot{\sim}} \mathrm{N}(0,1) \quad \Longrightarrow \quad \frac{(\hat{\theta}-\theta)^2}{\mathrm{var}(\hat{\theta})} \mathrel{\dot{\sim}} \chi^2_1$$

3. En el caso multivariado el estadístico $\hat{\boldsymbol{\theta}}$ tiene distribución asintótica normal.

$$\hat{\boldsymbol{\theta}} \mathrel{\dot{\sim}} \mathrm{N}(\boldsymbol{\theta}, V) \quad \Longrightarrow \quad (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})V^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \mathrel{\dot{\sim}} \chi^2_\nu$$

−ebz

Distribución muestral de *scores*

Considere la función de *score* $s(\boldsymbol{\theta}; y) = \partial\ell(\boldsymbol{\theta}; y)/\partial\boldsymbol{\theta}$. En particular, $s(\boldsymbol{\beta}) = \left(\frac{\partial\ell}{\partial\beta_1}, \ldots, \frac{\partial\ell}{\partial\beta_p}\right)'$

Y sabemos que

$$\begin{aligned} \mathbb{E}[s(\boldsymbol{\beta})] &= \mathbf{0} \\ \mathrm{var}(s(\boldsymbol{\beta})) &= \mathbb{E}[s(\boldsymbol{\beta})s(\boldsymbol{\beta})'] = \mathcal{I} \end{aligned}$$

Entonces, por el *teorema central de límite (TCL)*, $s(\boldsymbol{\beta}) \sim \mathrm{N}(\mathbf{0}, \mathcal{I})$, y por lo tanto

$$s(\boldsymbol{\beta})'\mathcal{I}^{-1}s(\boldsymbol{\beta}) \sim \chi^2_p$$

Ejemplo: Sea $y_1, \ldots, y_n$ una *m. a.* de $y \sim \mathrm{N}(\mu, \sigma^2)$, con σ^2 conocida. Entonces,

$$\ell(\mu; y) = -\frac{1}{2\sigma^2}\sum(y_i - \mu)^2 - \frac{n}{2}\log(2\pi\sigma^2) \tag{7}$$

$$s(\mu; y) = \frac{\partial\ell}{\partial\mu} = \frac{1}{\sigma^2}\sum(y_i - \mu) = \frac{n}{\sigma^2}(\bar{y} - \mu) \tag{8}$$

$$\mathcal{I} = \mathrm{var}(s(\mu)) = \frac{n^2}{\sigma^4}\left(\frac{\sigma^2}{n}\right) = \frac{n}{\sigma^2} \tag{9}$$

Por otro lado, $\frac{\bar{y}-\mu}{\sqrt{\sigma/n}} \sim \mathrm{N}(0, 1)$, por lo que efectivamente

$$s(\boldsymbol{\beta})'\mathcal{I}^{-1}s(\boldsymbol{\beta}) = \frac{\frac{n^2}{\sigma^4}(\bar{y} - \mu)^2}{\frac{n}{\sigma^2}} = \frac{n}{\sigma^2}(\bar{y} - \mu)^2 \sim \chi^2_1$$

−ebz

Distribución muestral de EMV $\hat{\beta}$

Sea $\hat{\beta}$ el EMV del vector de parámetros β, y suponga que $\hat{\beta}$ está cercano a β. Entonces, por aproximación de primer orden de Taylor a $s(\beta)$,

$$s(\beta) \approx s(\hat{\beta}) + H(\hat{\beta})(\beta - \hat{\beta})$$

donde $H(\hat{\beta}) = \left(\frac{\partial^2 s}{\partial \beta_j \partial \beta_k} \right) \Big|_{\beta=\hat{\beta}}$, y que por la *ley de los grandes números (LGN)* converge a $\mathbb{E}[H(\beta)] = -\mathcal{I}(\beta)$.
Luego, para grandes muestras, se tiene que

$$s(\beta) \approx s(\hat{\beta}) - \mathcal{I}(\beta)(\beta - \hat{\beta})$$

Pero $\hat{\beta}$ es EMV por lo que $s(\hat{\beta}) = 0$. Luego,

$$(\hat{\beta} - \beta) \approx \mathcal{I}^{-1}(\beta)s(\beta)$$

Ahora bien, $\mathbb{E}[(\hat{\beta} - \beta)] = 0$ pues $\mathbb{E}[s] = 0$. Entonces,

$$\text{var}(\hat{\beta} - \beta) = \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \mathcal{I}^{-1} E[ss']\mathcal{I}^{-1} = \mathcal{I}^{-1}$$

y por el *teorema del límite central*

$$(\hat{\beta} - \beta) \dot{\sim} \mathrm{N}(0, \mathcal{I}^{-1}) \tag{10}$$

y por lo tanto,

$$W = (\hat{\beta} - \beta)'\mathcal{I}(\hat{\beta} - \beta) \dot{\sim} \chi^2_p \tag{11}$$

El estadístico W se conoce como *estadístico de Wald*.

Nota: en el caso de los modelos lineales con respuestas normales las distribuciones de $(\hat{\beta} - \beta)$ y del estadístico de Wald son resultados exactos.

−ebz

Intervalos de confianza para los coeficientes β

La expresión (10) muestra la normalidad asintótica de $\hat{\beta}$. Luego, si se denota $V = \mathcal{I}^{-1}$, entonces

$$\hat{\beta} \mathrel{\dot{\sim}} \mathrm{N}(\beta, V)$$

Salvo el caso normal, los resultados distribucionales son asintóticos por lo que se tienen los siguientes resultados:

1. El *error estándar* de las estimaciones $\hat{\beta}_j$ es $\sqrt{v_{jj}}$. Esto es,
$$\mathrm{se}(\hat{\beta}_j) \approx \sqrt{v_{jj}}$$
2. Luego, intervalos del 95 % de confianza *aproximados* para los coeficientes $\hat{\beta}$ serían
$$\hat{\beta}_j \pm 1.96\sqrt{v_{jj}}$$
3. La correlación entre coeficientes están dados por
$$\mathrm{corr}(\hat{\beta}_j, \hat{\beta}_k) \approx \frac{v_{jk}}{\sqrt{v_{jj} v_{kk}}}$$

   En general la matriz de información $\mathcal{I}$ depende de β, por lo que en la práctica se emplea en su lugar su estimación $\hat{\beta}$. O bien, en ocasiones, en lugar de $\mathcal{I}$ se emplea $-H(\hat{\beta})$ como una estimación de la información.

–ebz

Ejemplo[6]

Sea $y_1, \ldots, y_n$ una *m. a.* donde $y_i \sim \mathrm{N}(x_i'\beta, \sigma^2)$, con σ^2 conocida. Sea X la matriz con filas de observaciones x_i' y tal que $X'X$ es no singular. En este caso,

$$\mathbb{E}[y_i] = \mu_i = x_i'\beta$$

La función liga es la *identidad*. Esto es, $\eta_i = g(\mu_i) = x_i'\beta$, por lo que $\partial\mu_i/\partial\eta_i = 1$.
Se tienen los siguientes resultados:

$$\mathcal{I}_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\mathrm{var}(y_i)} \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}x_{ik} \quad \Longrightarrow \quad \mathcal{I} = \frac{1}{\sigma^2}X'X$$

$$w_{ii} = \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 / \mathrm{var}(y_i) = \frac{1}{\sigma^2} \quad \Longrightarrow \quad W = \frac{1}{\sigma^2}I_n$$

$$z_i = \sum_j x_{ij}\hat{\beta}^{(r-1)} + (y_i - \mu_i)\left(\frac{\partial\mu_i}{\partial\eta_i}\right) \quad \Longrightarrow \quad z = X\hat{\beta} + y - X\hat{\beta} = y$$

$$X'WX\hat{\beta}^{(r)} = X'Wz \quad \Longrightarrow \quad X'X\hat{\beta} = X'y$$

y por lo tanto, resolviendo el sistema de *ecuaciones normales*

$$\hat{\beta} = (X'X)^{-1}X'y$$

[6] Dobson (1990)

Ejemplo (cont.)

Se sabe que $\hat{\beta} \sim \mathrm{N}(\beta, \sigma^2(X'X)^{-1})$. Entonces las siguientes distribuciones son exactas

$$(\hat{\beta} - \beta) \sim \mathrm{N}(0, \mathcal{I}^{-1})$$

con $\mathcal{I} = \frac{1}{\sigma^2}(X'X)$ y consecuentemente

$$(\hat{\beta} - \beta)'\mathcal{I}(\hat{\beta} - \beta) \sim \chi^2_p$$

−ebz

Bondad del ajuste

Para valorar lo adecuado del ajuste del *modelo de interés (MI)* se compara su verosimilitud con la correspondiente verosimilitud del *modelo maximal o saturado (MS)*.

El *modelo maximal* es aquel con la misma distribución y función liga que el modelo de interés y con tantos parámetros como observaciones, por lo que puede se supone que ofrece una *descripción completa* de los datos (dada la distribución).

Sean $L(\hat{\beta}_{\text{máx}}; y)$ y $L(\hat{\beta}; y)$ las funciones de verosimilitud de MS y de MI respectivamente, evaluados en sus correspondientes EMV. Si el modelo de interés describe adecuadamente los datos se puede esperar que $L(\hat{\beta}; y)$ y $L(\hat{\beta}_{\text{máx}}; y)$ estén cercanos. En caso contrario $L(\hat{\beta}; y)$ será mucho menor que $L(\hat{\beta}_{\text{máx}}; y)$, lo que sugiere el uso del estadístico del *cociente de verosimilitud generalizado*

$$\Lambda = \frac{L(\hat{\beta}_{\text{máx}}; y)}{L(\hat{\beta}; y)}$$

como una *medida de la bondad del ajuste*. Equivalentemente, tomando logaritmos,

$$\log \Lambda = \ell(\hat{\beta}_{\text{máx}}; y) - \ell(\hat{\beta}; y)$$

Distribución muestral del log-de-verosimilitud ℓ

Si se aproxima el log-de-verosimilitud por Taylor alrededor del vector de parámetros real β,

$$\ell(\beta; y) \approx \ell(\hat{\beta}; y) + (\beta - \hat{\beta})' s(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})' H(\hat{\beta})(\beta - \hat{\beta}) \tag{12}$$

donde $s(\hat{\beta}) = \frac{\partial \ell}{\partial \beta}\big|_{\beta=\hat{\beta}}$ y $H(\hat{\beta}) = \left(\frac{\partial^2 s}{\partial \beta_j \partial \beta_k}\right)\Big|_{\beta=\hat{\beta}}$. Por otro lado, como $\hat{\beta}$ es EMV $s(\hat{\beta}) = 0$, y por LGN, $H(\hat{\beta}) \to \mathbb{E}[H(\hat{\beta})] = -\mathcal{I}$. Entonces, reordenando (12)

$$\ell(\beta; y) - \ell(\hat{\beta}; y) \approx \frac{1}{2}(\hat{\beta} - \beta)' \mathcal{I} (\hat{\beta} - \beta)$$

Se sigue de la distribución del estadístico W de Wald (11) que

$$2 \log \Lambda = 2 \left[\ell(\beta; y) - \ell(\hat{\beta}; y) \right] \mathrel{\dot{\sim}} \chi^2_p$$

Estadísticos de prueba basados en el estadístico anterior son empleados para valorar la bondad del ajuste de un modelo y la comparación entre modelos.

−ebz

Devianza D

Nelder y Wedderburn (1972) definieron el estadístico *log del cociente de la verosimilitud* como la *devianza D* (escalada)

$$D = 2\log\Lambda = 2\left[\ell(\hat{\beta}_{\text{máx}}\,;y) - \ell(\hat{\beta};y)\right]$$

D puede descomponerse como

$$D = \left\{ \underbrace{\left[\ell(\hat{\beta}_{\text{máx}}\,;y) - \ell(\beta_{\text{máx}}\,;y)\right]}_{\chi^2_n} - \underbrace{\left[\ell(\hat{\beta};y) - \ell(\beta;y)\right]}_{\chi^2_p} + \underbrace{\left[\ell(\beta_{\text{máx}}\,;y) - \ell(\beta;y)\right]}_{\geq 0} \right\}$$

En grandes rasgos, si los primeros dos sumandos son independientes y el tercero es cercano a cero, entonces

$$D \dot{\sim} \chi^2_{n-p}$$

si el modelo es adecuado. Si por el contrario el modelo no es bueno el tercer término será grande y D será mucho mayor que lo esperado por una distribución χ^2_{n-p}.
En la práctica uno tiende a comparar el D calculado de los datos con $(n-p)$, el valor medio de la distribución.
Nota: en general la descomposición anterior es una aproximación para la distribución muestral del estadístico aunque para el caso normal el resultado es exacto.

Ejemplo

Considere $y_1, \ldots, y_n$ *m. a.* con $y_i \sim \mathrm{N}(\mu_i, \sigma^2)$, σ^2 común y conocida. Entonces, la función *log-de-verosimilitud* está dada por

$$\ell(\beta; y) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i)^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

En el caso del *modelo maximal* (MS), $\mathbb{E}[y_i] = \mu_i$, así $\beta = (\mu_i, \ldots, \mu_n)'$. Luego, el EMV $\hat{\beta}_{\text{máx}}$, queda como $\hat{\mu}_i = y_i$, para $i = 1, \ldots, n$, y por lo tanto

$$\ell(\hat{\beta}_{\text{máx}}; y) = -\frac{n}{2} \log(2\pi\sigma^2)$$

Considere ahora el *modelo reducido* (MR) que tiene un único parámetro común μ. Entonces, $\hat{\mu} = \bar{y}$ y el correspondiente log-de-verosimilitud

$$\ell(\hat{\beta}; y) = -\frac{1}{2\sigma^2} \sum_{i=-1}^{n} (y_i - \bar{y})^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

−ebz

Ejemplo (cont.)

Por lo tanto, la devianza D queda dada por

$$D = 2\left[\ell(\hat{\beta}_{\text{máx}}\,; y) - \ell(\hat{\beta}; y)\right] = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n-1}{\sigma^2}\,s^2$$

Si el modelo de un solo parámetro es correcto, sabemos que $s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$ y que $\frac{n-1}{\sigma^2}s^2 \sim \chi^2_{n-1}$, por lo que

$$D \sim \chi^2_{n-1}$$

es un resultado exacto.

Inferencia – Pruebas de hipótesis

Contrastes de hipótesis sobre el vector de parámetros β se llevan a cabo utilizando las distribuciones asintóticas de los siguientes estadísticos:

$$(\hat{\beta} - \beta) \dot{\sim} \mathrm{N}(0, \mathcal{I}^{-1}) \qquad \text{EMV}$$

$$(\hat{\beta} - \beta)' \mathcal{I} (\hat{\beta} - \beta) \dot{\sim} \chi^2_p \qquad \text{Wald}$$

$$s(\beta)' \mathcal{I}^{-1} s(\beta) \dot{\sim} \chi^2_p \qquad \text{score}$$

Alternativamente se puede utilizar la *diferencias de devianzas* a la manera de *suma extra de cuadrados*. Para esto considere los modelos reducido (MR) y completo (MC) tales que las hipótesis se pueden plantear como

$$H_0 : \beta = B_0 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \qquad vs. \qquad H_1 : \beta = B_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

con $1 \leq q < p$.

Inferencia – Pruebas de hipótesis

Ahora, utilizando el estadístico $D = 2\log\Lambda$,

$$\Delta D = D_0 - D_1 = 2\left[\ell(\hat{B}_1; y) - \ell(\hat{B}_0; y)\right]$$

Si ambos modelos son aceptables, $\Delta D \dot{\sim} \chi^2_{q-p}$, siempre que ciertas condiciones de independencia se cumplan. Si ΔD es grande y por lo mismo su correspondiente valor$-p$ es pequeño, se decide en favor del modelo completo. En caso contrario se decide en favor de la hipótesis H_0, correspondiente a un modelo *aceptable* más simple o parsimonioso.

Ejemplo[7]

En el caso normal el incremento de devianzas da lugar exactamente a la prueba F, empleada en las pruebas de hipótesis por medio de la *suma extra de cuadrados*.
A saber, sean $\hat{\mu}_i^0$ y $\hat{\mu}_i^1$ las respuestas ajustadas bajo los modelos reducido (H_0) y completo (H_1) respectivamente. Entonces,

$$D_k = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu}_i^k)^2, \qquad k = 0, 1$$

Si H_1 es válida, $D_1 \dot{\sim} \chi^2_{n-p}$. Y si H_0 es también correcta $D_0 \dot{\sim} \chi^2_{n-q}$ y por lo tanto, $\Delta D = D_0 - D_1 \dot{\sim} \chi^2_{p-q}$.
Ahora bien, para eliminar la dependencia del parámetro σ^2, sea

$$F = \frac{\frac{D_0 - D_1}{p-q}}{\frac{D_1}{n-p}} = \frac{\left[\sum(y_i - \hat{\mu}_i^0)^2 - \sum(y_i - \hat{\mu}_i^1)^2\right]/(p-q)}{\sum(y_i - \hat{\mu}_i^1)^2/(n-p)}$$

que en el caso normal se tiene la distribución exacta $F \sim F_{p-q,n-p}$, como se vio que era el caso de la *suma extra de cuadrados*.

[7] Dobson (1990)

Residuales

En el caso normal se definen los *residuales* r_i por

$$y_i = \hat{\mu}_i + (y_i - \hat{\mu}_i) = \hat{\mu}_i + r_i$$

Esta definición no es empleada en MLG pues en general la varianza $\mathrm{var}(y_i)$ depende del nivel de la respuesta.

Residual	Definición	Observaciones
Normal	$r_i = y_i - \hat{\mu}_i$	Impráctico pues en MLG la varianza no es constante.
Intuitivo	$r_{Ri} = y_i - g^{-1}(x_i'\hat{\beta})$	Impráctico pues en MLG la varianza no es constante.
Práctico	$r_{Wi} = (y_i - \hat{\mu}_i)\left.\frac{\partial \eta}{\partial \mu}\right\|_{\mu=\hat{\mu}}$	Utilizando el resultado de la última iteración.
Pearson	$r_{Pi} = \dfrac{y_i - \hat{\mu}_i}{\sqrt{\mathrm{var}(y_i)}}$	Residual estandarizado.
Anscombe	$r_{Ai} = A(y_i)$	La función $A(\cdot)$ se elige dependiendo de la distribución de y_i de manera tal que $A(y_i)$ sea "tan normal" como sea posible.
Devianza	$r_{Di} = \mathrm{sign}(y_i - \hat{\mu}_i)\sqrt{d_i}$	Residual de la devianza donde los d_i son los sumandos de la devianza. Esto es, si $D = \sum d_i$, entonces, $D = \sum r_{Di}^2$.

Nota: El análisis de residuales en MLG es similar al caso normal. Se espera un comportamiento aleatorio de los residuales. Esto es, sin patrones aparentes.

Los residuales más populares son los de Pearson y los de devianza.

Residuales

Con relación a los residuales de Anscombe, la función A es tal que $A(\mu) = \int_{-\infty}^{\mu} v^{-1/3}(t)dt$, con $v(t)$ la varianza función de la media. Luego, el residual de Anscombe está definido como

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{A(\hat{\mu}_i)\sqrt{\text{var}A(\hat{\mu}_i)}}$$

Por lo que se tienen los siguientes casos específicos:

$$\begin{array}{rccl}
\text{Normal :} & r_{Ai} & = & y_i - \hat{\mu}_i \\
\text{Gaussiana inversa :} & r_{Ai} & = & \left(\log(y_i) - \log(\hat{\mu}_i)\right) / (\hat{\mu}_i)^{1/2} \\
\text{Gamma} & r_{Ai} & = & 3((y_i - \hat{\mu}_i)^{1/3} - 1) \\
\text{Poisson} & r_{Ai} & = & \frac{3}{2}((y_i^{2/3} - \hat{\mu}_i)^{1/6} - \hat{\mu}^{1/2}) \\
\text{Binomial} & r_{Ai} & = & \sqrt{m_i}\left(B(y_i, \frac{2}{3}, \frac{2}{3}) - B(\hat{\mu}_i, \frac{2}{3}, \frac{2}{3})\right)\left(\hat{\mu}_i(1 - \hat{\mu}_i)\right)^{1/6}
\end{array}$$

donde B denota la *función beta*, $B(y, a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1}dt$.
Los correspondientes residuales de Anscombe *estandarizados* (r_{Asi}) y *studentizados* (r_{Ati}) dados por

$$r_{Asi} = \frac{r_{Ai}}{\sqrt{\hat{\phi}(1 - h_{ii})}}, \qquad \text{y} \qquad r_{Ati} = \frac{r_{Ai}}{\sqrt{\hat{\phi}_{(i)}(1 - h_{ii})}}$$

con $\hat{\phi}$ la estimación del parámetro de dispersión. Véase ecuación (2).

Regresión Logística

Regresión Logística

Respuesta binaria

Considere una variable que puede tomar dos valores solamente: *éxito* ó *fracaso*; *sí* ó *no*; 1 ó 0. Así,

$$Z = \begin{cases} 1 & \text{éxito, sí} \\ 0 & \text{fracaso, no} \end{cases}$$

y tal que

$$P(Z=1)=\pi, \qquad P(Z=0)=1-\pi, \qquad f(z)=\pi^{z}(1-\pi)^{1-z}$$

Z se dice que sigue una *distribución Bernoulli parámetro* π. Sea $Z_1, \ldots, Z_r$, independientes con $Z_j \sim \mathrm{Ber}(\pi_j)$, entonces

$$f(\boldsymbol{z};\pi)=\prod_{j=1}^{r} f(z_j;\pi)=\prod \pi_j^{z_j}(1-\pi_j)^{1-z_j}$$

que se puede reescribir como

$$\begin{aligned} f(\boldsymbol{z};\pi) &= \exp\left\{\sum_{j=1}^{r} z_j \log(\pi_j)+\sum_{j=1}^{r}(1-z_j)\log(1-\pi_j)\right\} \\ &= \exp\left\{\sum z_j \log\frac{\pi_j}{1-\pi_j}+\sum \log(1-\pi_j)\right\} \\ &= \exp\{a(z)b(\theta)+c(\theta)+d(z)\} \end{aligned}$$

Esto es, *la distribución Bernoulli es miembro de la familia exponencial.*

Si las *v. a.* Z_j son además idénticamente distribuidas con $\pi_j = \pi$, entonces $Y = \sum_{j=1}^{r} Z_j \sim \text{Bin}(r, \pi)$.
La *v. a.* Y representa el número de éxitos en la muestra de tamaño r. La distribución tiene el soporte $R_Y = \{0, 1, \ldots, r\}$ y *f. d. p.*

$$f(y; \pi) = \binom{r}{y} \pi^y (1-\pi)^{r-y} I_{R_Y}(y)$$

Supongamos una muestra de tamaño n donde $Y_i \sim \text{Bin}(n_i, \pi_i)$. Entonces,

$$\ell(\boldsymbol{\pi}; \boldsymbol{y}) = \sum_{i=1}^{n} \left[y_i \log \frac{\pi_i}{1-\pi} + n_i \log(1-\pi_i) + \log \binom{n_i}{y_i} \right]$$

La distribución binomial también es miembro de la familia exponencial.

Modelos lineales generalizados

Se desea describir las proporciones de éxito $P_i = Y_i/n_i$ en términos de *factores* (variables cualitativas) y *covariables* (cuantitativas y continuas). Para esto,

$$g(\pi_i) = x_i' \beta$$

donde las x son las *variables explicativas*, β el *vector de parámetros* y g la *función liga*. El caso más simple es el *modelo lineal*

$$\pi = x'\beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ip}$$

donde la función liga es la identidad. El problema con tal modelo es que puede producir valores de probabilidad π fuera del intervalo $[0, 1]$.
Para asegurar que el valor de π esta en el intervalo se usan las funciones de probabilidad acumulada

$$\pi = g^{-1}(x'\beta) = \int_{-\infty}^{u} f(t)dt$$

y donde f es una función de densidad de probabilidad (*f. d. p.*) llamada *distribución de tolerancia*.
[Si f es *f. d. p.* entonces, $f(t) \geq 0$, y $\int f(t)dt = 1$.]

Modelos dosis-respuesta

Los primeros modelos tipo regresión lineal usados para ajustar datos binomiales fue en bioensayos. Respuestas como proporción de animales que sobreviven determinada dosis de toxinas. Tales respuestas son llamadas *respuestas cuantiles*.

Modelo probit Si la distribución de tolerancia es la normal,

$$\pi = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt$$

donde Φ es la *f. p. a.* de la normal estándar.Luego,

$$\Phi^{-1}(\pi) = \beta_0 + \beta_1 x$$

con $\beta_0 = -\mu/\sigma$ y $\beta_1 = 1/\sigma$ y la función liga es la inversa de la *f. p. a.* Φ^{-1}.

Los *modelos probit* se usan en áreas de las ciencias biológicas y sociales donde se dan interpretaciones naturales del modelo. Por ejemplo, $x = \mu$ es llamada al *dosis letal mediana (LD(50))*.

Modelo logístico o *logit*

Modelo que permite resultados similares al modelo *probit* pero computacionalmente más sencillo. Para este caso la distribución de tolerancia es

$$f(t) = \frac{\beta_0 \exp(\beta_0 + \beta_1 t)}{[1 + \exp(\beta_0 + \beta_1 t)]^2}$$

Lo que implica que las probabilidades π quedan determinadas por

$$\pi = \int_{-\infty}^{x} f(t)dt = \frac{\exp(\beta_0 + \beta_1 t)}{1 + \exp(\beta_0 + \beta_1 t)} = g^{-1}(x'\beta)$$

O bien,

$$\eta = g(\mu) = g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$$

La función liga $g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$ se conoce como *función logística*.
El comportamiento de $f(t)$ y $\pi(x)$ es muy parecido al de la función *probit* excepto en las colas de las distribuciones.

Momios

El cociente $\frac{\pi}{1-\pi}$ se conoce como *momios* o *ventaja* (*odds* en inglés).

Considere la tabla de contingencia de la derecha de gente expuesta o no a cierto contaminante después de cierto tiempo y si se ha enfermado o no.

	enfermo	*no enfermo*
expuesto	π_1	$1-\pi_1$
no expuesto	π_2	$1-\pi_2$

con $0 < \pi_i < 1$

Los momios de la enfermedad, para ambos grupos, expuestos o no, son

$$\theta_i = \frac{\pi_i}{1-\pi_i}, \quad i = 1,2 = \textit{expuesto / no-expuesto}$$

Ejemplo:

probabilidad π	ventaja $\frac{\pi}{1-\pi}$	momios	*logit* $\log\left(\frac{\pi}{1-\pi}\right)$	explicación
$1/2$	1.00	$1:1$	0	éxito o fracaso indistinto
$3/4$	3.00	$3:1$	>0	éxito más posible que fracaso
$1/5$	0.25	$1:4$	<0	éxito menos posible que fracaso

Considere el ejemplo de la tabla de contingencia de la derecha de gente expuesta o no a cierto contaminante después de cierto tiempo y si se ha enfermado o no.

	enfermo	*no enfermo*
expuesto	π_1	$1-\pi_1$
no expuesto	π_2	$1-\pi_2$

con $0 < \pi_i < 1$

Los momios de la enfermedad, para ambos grupos, expuestos o no, son

$$\theta_i = \frac{\pi_i}{1-\pi_i}, \quad i = 1, 2 = \textit{expuesto / no-expuesto}$$

Entonces, la *razón de momios (o ventajas)* es

$$\phi = \frac{\theta_1}{\theta_2} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$$

que es la medida de la posibilidad relativa de la enfermedad para el grupo de *expuestos* sobre los *no-expuestos*.

Para un modelo logístico simple (un solo regresor): $\log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x$. Luego,

$$\frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1 x} = e^{x'\beta} \quad \Longrightarrow \quad \pi_i = \frac{e^{x'\beta}}{1+e^{x'\beta}}$$

Si

$$\phi = 1 = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \quad \Longrightarrow \quad \frac{\pi_1}{1-\pi_1} = \frac{\pi_2}{1-\pi_2}$$

lo que significa que el *riesgo* de la enfermedad (proporción de enfermos) es la misma para los grupos de *expuestos* o *no-expuestos*.
Se conoce como la *diferencia logit* a

$$\log\phi = \log\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} = g(\pi_1) - g(\pi_2) = \beta_1$$

cuya interpretación es una de las razones fundamentales de porque la regresión logística ha mostrado ser de gran importancia práctica.

Si

$$\phi = 1 = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \quad \Longrightarrow \quad \frac{\pi_1}{1-\pi_1} = \frac{\pi_2}{1-\pi_2}$$

lo que significa que el *riesgo* de la enfermedad (proporción de enfermos) es la misma para los grupos de *expuestos* o *no-expuestos*.

Considere ahora J tablas de contingencia 2×2, una para nivel de un factor, por ejemplo, grupo de edad: $j = 1, \ldots, J$. Entonces,

$$\pi_{ij} = \frac{\exp(\beta_{0i} + \beta_{1i} x_i)}{1 + \exp(\beta_{0i} + \beta_{1i} x_i)}, \quad i = 1, 2;\ j = 1, \ldots, J$$

Si $\beta_{11} = \beta_{12}$ constante sobre las J tablas, entonces $\phi =$ constante y

$$\begin{aligned}
\log \phi &= \log \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \\
&= \log \frac{\pi_1}{1-\pi_1} - \log \frac{\pi_2}{1-\pi_2} \\
&= (\beta_{01} - \beta_{11} x_j) - (\beta_{02} - \beta_{12} x_j) \\
&= (\beta_{01} - \beta_{02}) - (\beta_{11} - \beta_{12}) x_j \\
&= (\beta_{01} - \beta_{02})
\end{aligned}$$

−ebz

Modelo de *valor extremo*

Si se emplea ahora la *distribución de valor extremo*

$$f(t) = \beta_1 \exp\left\{(\beta_0 + \beta_1 t) - \exp(\beta_0 + \beta_1)\right\}$$

como función de tolerancia las probabilidades quedan dadas por:

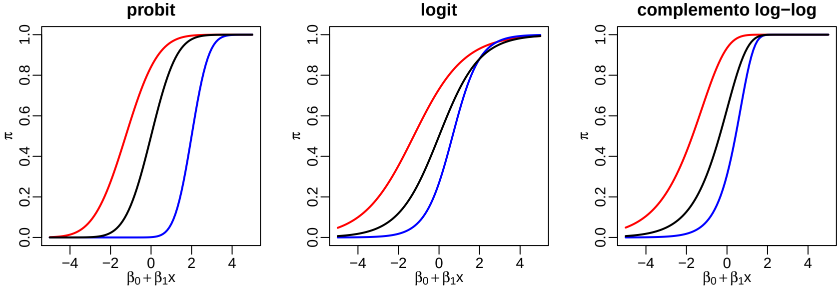
$$\pi = 1 - \exp\{-\exp(\beta_0 + \beta_1 x)\}$$

y por lo tanto,

$$\eta = g(\pi) = \log\left[-\log(1 - \pi)\right] = \beta_0 + \beta_1 x$$

En este caso la función liga es $\log\left[-\log(1-\pi)\right]$, llamada *función complementaria log-log*.
Los modelos de valor extremos son empleados en áreas como el *análisis de sobrevivencia* o *confiabilidad*.
La distribución es similar a los modelos *probit* o *logit* al rededor de 0.5 y haciéndose la diferencia más clara cerca del 0 o 1.

Funciones liga para respuestas binarias



Funciones liga

Probit: $\pi = \Phi^{-1}(x'\beta)$

Logit: $\pi = \dfrac{e^{x'\beta}}{1 + e^{x'\beta}}$

Complemento log-log: $\pi = 1 - \exp\{-\exp(x'\beta)\}$

Coeficientes:

$\beta_0 = 0$; $\beta_1 = 1$

$\beta_0 = 1$; $\beta_1 = 4/5$

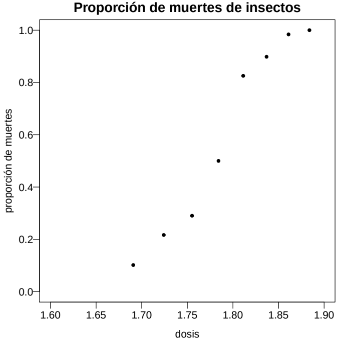
$\beta_0 = -3$; $\beta_1 = 3/2$

−ebz

Problema de mortalidad de insectos[8]

La siguiente tabla muestra el número de insectos muertos después de 5 horas de exposición a gas carbónico a distintas concentraciones. La gráfica presenta la proporción de insectos muertos (y_i/n_i) como función de la dosis.

dosis x_i	número de insectos n_i	número de muertes y_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60



[8] Dobson (1990)

Problema de mortalidad de insectos (cont.)

El modelo logístico sería, para $i = 1, \ldots, n$

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

o bien,

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_i$$

En este caso, el estadístico *log-razón-de-verosimilitud* es

$$D = \sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n - y_i}{n - \hat{y}_i}\right)\right]$$

Modelo ***logit:***

```
Call:
glm(formula = y ~ x, family = binomial("logit"), weights = dat$n)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16
x             34.270      2.912   11.77   <2e-16

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom (valor-p=0.082)
AIC: 41.43
Number of Fisher Scoring iterations: 4
```

–ebz

Problema de mortalidad de insectos (cont.)

Modelo ***probit:***

```
Call:
glm(formula = y ~ x, family = binomial("probit"), weights = dat$n)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -34.935      2.648  -13.19   <2e-16
x             19.728      1.487   13.27   <2e-16

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  10.120  on 6  degrees of freedom  (valor-p=0.120)
AIC: 40.318
Number of Fisher Scoring iterations: 4
```

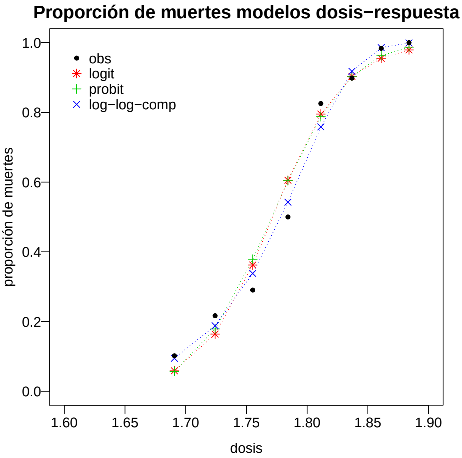
Modelo ***valor extremo:***

```
Call:
glm(formula = y ~ x, family = binomial("cloglog"), weights = dat$n)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -39.572      3.240  -12.21   <2e-16
x             22.041      1.799   12.25   <2e-16

    Null deviance: 284.2024  on 7  degrees of freedom
Residual deviance:   3.4464  on 6  degrees of freedom (valor-p=0.752)
AIC: 33.644
Number of Fisher Scoring iterations: 4
```

−ebz

Problema de mortalidad de insectos (cont.)

dosis x_i	número de insectos n_i	número de muertes y_i	predicciones de modelos logístico	probit	valor extremo
1.6907	59	6	3.46	3.36	5.59
1.7242	60	13	9.84	10.72	11.28
1.7552	62	18	22.45	23.48	20.95
1.7842	56	28	33.90	33.82	30.37
1.8113	63	52	50.10	49.62	47.78
1.8369	59	53	53.29	53.32	54.14
1.8610	62	61	59.22	59.66	61.11
1.8839	60	60	58.74	59.23	59.95
Devianza D			11.23	10.12	3.45



Regresión logística

Considere el modelo de regresión logística múltiple, para $i = 1, \ldots, n$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ip}$$

o bien,

$$\eta_i = g(\pi_i) = x_i'\beta, \qquad i = 1, \ldots, n$$

Los estimadores (EMV) del vector de parámetros β y consecuentemente de las probabilidades π tales que $g(\hat{\pi}) = x'\hat{\beta}$ se obtienen maximizando la función log-verosimilitud

$$\ell(\beta; y) = \sum_{i=1}^{n}\left[y_i \log(\pi_i) + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}\right]$$

que acepta aún casos con $n_i = 1$ y/o $y_i = 0$, que no los métodos de mínimos cuadrados.
El correspondiente estadístico razón log-verosimilitud está dado por

$$D = 2\left(\ell(\hat{\pi}_{\text{máx}}; y) - \ell(\hat{\pi}; y)\right)$$

donde $\hat{\pi}_{\text{máx}}$ es el vector EMV del *modelo maximal* y $\hat{\pi}$ el EMV del *modelo de interés*.

Para el *modelo maximal* se tiene que

$$\frac{\partial \ell}{\partial \pi_i} = \frac{y_i}{\pi_i} - \frac{n_i - y_i}{1 - \pi_i}$$

por lo que $\partial \ell / \partial \pi_i = 0 \Longrightarrow \hat{\pi}_{\text{máx},i} = y_i / n_i$. Así,

$$\ell(\hat{\pi}_{\text{máx}}; y) = \sum_{i=1}^{n} \left[y_i \log \left(\frac{y_i}{n_i} \right) + (n_i - y_i) \log \left(1 - \frac{y_i}{n_i} \right) + \log \binom{n_i}{y_i} \right]$$

Luego,

$$D = 2 \sum_{i=1}^{n} \left[y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n - y_i}{n_i - n_i \hat{\pi}_i} \right) \right]$$

Entonces, D tiene la forma

$$D = 2 \sum o \log \frac{o}{e}$$

donde o denota las frecuencias observadas y_i y $(n_i - y_i)$ de las celdas de una tabla de contingencia y e corresponde a las frecuencias estimadas $n_i \hat{\pi}_i$ y $(n - n_i \hat{\pi}_i)$, respectivamente.
D no depende de otros parámetros (como σ^2 de la normal) por lo que se puede usar como medida de bondad de ajuste y para pruebas de hipótesis usando la aproximación

$$D \sim \chi^2_q$$

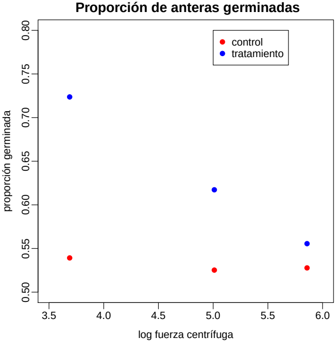
donde $q = 1 + p$, el numero de coeficientes β estimados.

–ebz

Problema de anteras embriogénicas de la planta de especie *Datura innoxia*[9]

La siguiente tabla muestra el número de anteras embriogénicas y_{jk} de planta de especie *Datura innoxia* obtenidas cuando un número de anteras n_{jk} fueron preparadas bajo distintas condiciones. Hay un *factor* cualitativo: un tratamiento de almacenaje a 3°C por 48 hrs. o bien bajo almacenamiento controlado; y una *covariable* continua con 3 valores de fuerza centrífuga. Interesa comparar el efecto del tratamiento de almacenaje después de ajustar, si es necesario, por la fuerza centrífuga.

Condición almacenaje		Fuerza centrífuga (g.) 40	150	350
Control	y_{ik}	55	52	57
	n_{1k}	102	99	108
Tratamiento	y_{2k}	55	50	50
	n_{2k}	76	81	90



[9]Dobson (1990)

Problema de anteras (cont.)

Ajuste de modelos por máxima verosimilitud

Modelo 1:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \underset{(0.628)}{0.239} + \underset{(0.998)}{1.977 I_{\text{Trt}}} - \underset{(0.127)}{0.023 \log(x)} - \underset{(0.199)}{0.319 \log(x) * I_{\text{Trt}}} \qquad \begin{aligned} D_1 &= 0.028 \\ AIC &= 37.596 \end{aligned}$$

Modelo 2:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \underset{(0.487)}{0.877} + \underset{(0.175)}{0.407 I_{\text{Trt}}} - \underset{(0.097)}{0.155 \log(x)} \qquad \begin{aligned} D_2 &= 2.619 \\ AIC &= 38.187 \end{aligned}$$

Modelo 3:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \underset{(0.148)}{1.021} + \quad - \underset{(0.097)}{0.148 \log(x)} \qquad \begin{aligned} D_3 &= 8.092 \\ AIC &= 41.66 \end{aligned}$$

−ebz

Problema de anteras (cont.)

Código:

```
> logx <- log(dat$x)
> dat$prop <- dat$y/dat$n
> dat$logx <- log(dat$x)
> print(dat)

> mod1 <- glm(y~log*f,
family=binomial("logit"),data=dat,weights=n)
> print(summary(mod1)
> print(anova(mod1))
```

Datos:

```
f    x   n  y     logx         p
ctrl  40 102 55 3.688879 0.5392157
ctrl 150  99 52 5.010635 0.5252525
ctrl 350 108 57 5.857933 0.5277778
 trt  40  76 55 3.688879 0.7236842
 trt 150  81 50 5.010635 0.6172840
 trt 350  90 50 5.857933 0.5555556
```

Problema de anteras (cont.)

Salida:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.23389    0.62839   0.372   0.7097
logx        -0.02274    0.12685  -0.179   0.8577
ftrt         1.97711    0.99802   1.981   0.0476
logx:ftrt   -0.31862    0.19888  -1.602   0.1091

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 10.451974  on 5  degrees of freedom
Residual deviance:  0.027728  on 2  degrees of freedom
AIC: 37.596
Number of Fisher Scoring iterations: 3


Analysis of Deviance Table
Model: binomial, link: logit
Response: y
Terms added sequentially (first to last)
       Df Deviance Resid. Df Resid. Dev
NULL                      5    10.4520
logx    1   2.3604        4     8.0916
f       1   5.4727        3     2.6188
logx:f  1   2.5911        2     0.0277
```

–ebz

Problema de anteras (cont.)

Considere el modelo general:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 I_{\mathrm{Trt}} + \beta_2 \log(x) + \beta_3 \log(x) * I_{\mathrm{Trt}}$$

• Para probar la igualdad de pendientes de ambos grupos de almacenaje:

$$H_0 : \beta_3 = 0 \textit{ vs. } H_1 : \beta_3 \neq 0$$

se sigue que $D_2 - D_1 \sim \chi_1^2$. En este caso,

$$D_2 - D_1 = 2.619 - 0.028 = 2.591 \qquad (\text{valor-}p = 0.1074)$$

por lo que los datos no apoyan fuertemente la diferencia de pendientes del término interacción.
• Por otro lado, para probar la significancia del efecto del tratamiento de almacenaje, *después de ajustar por la covariable (log de) fuerza centrífuga*:

$$H_0 : \beta_1 = 0 \textit{ vs. } H_1 : \beta_1 \neq 0$$

$$D_3 - D_2 = 8.092 - 2.619 = 5.473 \qquad (\text{valor-}p = 0.0193)$$

En este caso, los datos sugieren que sí hay un efecto del tratamiento del almacenaje en la proporción de anteras germinadas.

−ebz

Problema de anteras (cont.)

Condición de almacenaje	fuerza x	$\log(x)$	proporciones y	Modelos $\hat{y}$ 1	2	3
Control	40	3.689	0.539	0.537	0.576	0.617
	150	5.011	0.525	0.530	0.526	0.570
	350	5.858	0.528	0.525	0.493	0.539
Tratamiento	40	3.689	0.724	0.721	0.671	0.617
	150	5.011	0.617	0.623	0.625	0.570
	350	5.858	0.556	0.553	0.593	0.539
		D		0.028	2.619	8.092

La tabla muestra las proporciones ajustadas por los 3 modelos. Note lo bien que el modelo 1 ajusta los datos. Ésto sucede porque se estiman *¡4 parámetros con base a 6 observaciones!*

Problema de recién nacidos bajos de peso [10]

El problema de niños con bajo peso al nacer es una condición que preocupa a los médicos pues constituye un grupo de alto riesgo para los problemas de mortalidad infantil y tasas de defectos al nacer. El comportamiento de la madre durante el embarazo (dieta, hábitos de fumar, cuidado prenatal, etc.) puede afectar de manera importante el tiempo de gestación y en consecuencia el peso normal de un niño al nacer.

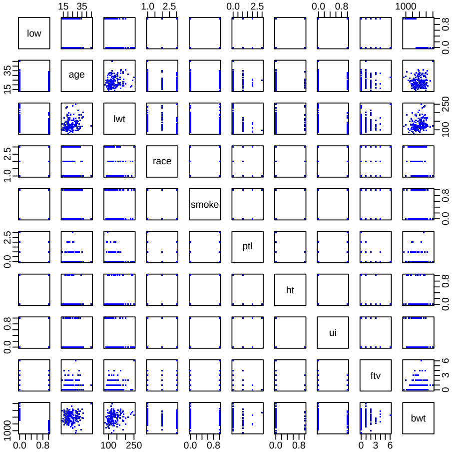
variable	código	etiqueta
Código de identificación:	...	`id`
Bajo peso:	0 = peso $\geq$ 2500 gr. 1 = peso $<$ 2500 gr.	`low`
Edad de la madre:	años	`age`
Peso antes de la última menstruación:	libras	`lwt`
Raza:	1=blanca;2=negra;3=otra	`race`
Hábito del fumar durante embarazo:	1=sí;0=no	`smoke`
Historia de labor prematura	0=ninguna;1=una;etc.	`ptl`
Historia de hipertensión	1=sí;0=no	`ht`
Irritación uterina:	1=sí;0=no	`ui`
Número de visitas médicas durante el primer trimestre:	0=ninguna;1=una;etc.	`ftv`
Peso al nacer:	gramos	`bwt`

```
===============================================
obs id low age lwt race smoke ptl ht ui ftv  bwt
-----------------------------------------------
  1 85   0  19 182    2     0   0  0  1   0 2523
  2 86   0  33 155    3     0   0  0  0   3 2551
  3 87   0  20 105    1     1   0  0  0   1 2557
  4 88   0  21 108    1     1   0  0  1   2 2594
  .  .   .   .   .    .     .   .  .  .   .    .
===============================================
```

[10] Hosmer and Lemeshow (1989); Venables and Ripley (2002).

Problema de recién nacidos bajos de peso (cont.)

Despliegue de datos



Problema de recién nacidos bajos de peso (cont.)

Manipulación de datos

Se considera la variable `low` como respuesta binaria y será modelada mediante una regresión logística.

No se espera que la variable `ptl` influya linealmente sobre la respuesta. Además son pocos los casos mayores que 1, por lo que se le trata como una variable binaria. Igual es el caso de la variable `ftv` que se reduce a tres categorías.

Código:

```
> options(contrasts=c("contr.treatment","contr.poly"))
> attach(birthwt)
> race <- factor(race,labels=c("white","black","others"))
> print(table(ptl))

> lwd <- factor(lwt < 110)
> ptd <- factor(ptl > 0)
> print(table(ftv))

> ftv <- factor(ftv)
> levels(ftv)[-(1:2)] <- "2+"
> print(table(ftv))

> bwt <- data.frame(low=factor(low),age,lwt,race,
            smoke=(smoke>0),ptd,ht=(ht>0),ui=(ui>0),ftv)
> detach()
> rm(race,ptd,ftv)
```

Salida:

```
ptl
  0   1   2   3
159  24   5   1

ftv
  0   1   2   3   4   6
100  47  30   7   4   1

ftv
  0   1  2+
100  47  42
```

–ebz

Problema de recién nacidos bajos de peso (cont.)

Primer ajuste de regresión logística

```
> birthwt.glm <- glm(low ~ age+lwt+race+smoke+ptd+ht+ui+ftv,
+                          binomial,dat=bwt)
> print(summary(birthwt.glm,cor=FALSE))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.82302    1.24471   0.661  0.50848
age         -0.03723    0.03870  -0.962  0.33602
lwt         -0.01565    0.00708  -2.211  0.02705
raceblack    1.19241    0.53597   2.225  0.02609
raceothers   0.74069    0.46174   1.604  0.10869
smokeTRUE    0.75553    0.42502   1.778  0.07546
ptdTRUE      1.34376    0.48062   2.796  0.00518
htTRUE       1.91317    0.72074   2.654  0.00794
uiTRUE       0.68019    0.46434   1.465  0.14296
ftv1        -0.43638    0.47939  -0.910  0.36268
ftv2+        0.17901    0.45638   0.392  0.69488
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 195.48  on 178  degrees of freedom
AIC: 217.48
Number of Fisher Scoring iterations: 4
```

Puesto que la respuesta es binaria, aunque el modelo sea correcto no se puede garantizar que la devianza siga aún aproximadamente una distribución χ^2, pero que tenga una valor parecido a los grados de libertad no sugiere problemas de ajuste.

–ebz

Problema de recién nacidos bajos de peso (cont.)

Selección de modelo 1:
Para la selección del modelo se recurre a la estrategia *stepwise* pero basada en cambios del estadístico AIC, el *criterio de información de Akaike*, dado por

$$\text{AIC} = -2\ell(\hat{\theta}; y) + 2q$$

Ajuste:

```
> birthwt.step <- stepAIC(birthwt.glm,trace=FALSE)
> print(birthwt.step$anova)

 Stepwise Model Path
 Analysis of Deviance Table

 Initial Model:
 low ~ age + lwt + race + smoke + ptd + ht + ui + ftv

 Final Model:
 low ~ lwt + race + smoke + ptd + ht + ui


    Step Df Deviance Resid. Df Resid. Dev      AIC
 1                         178   195.4755 217.4755
 2 - ftv  2 1.358185       180   196.8337 214.8337
 3 - age  1 1.017866       181   197.8516 213.851
```

Análisis de devianzas:

```
> print(anova(birthwt.step))

Analysis of Deviance Table
Model: binomial, link: logit
Response: low
Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev
NULL                    188     234.67
lwt    1   5.9813       187     228.69
race   2   5.4316       185     223.26
smoke  1   8.2444       184     215.01
ptd    1   7.9752       183     207.04
ht     1   6.5572       182     200.48
ui     1   2.6307       181     197.85
```

−ebz

Problema de recién nacidos bajos de peso (cont.)

Selección de modelo 2:

En esta segunda búsqueda se incluyen términos cuadráticos (escalados) de edad y peso de la madre.

Ajuste:

```
> birthwt.step2 <- stepAIC(birthwt.glm, ~ .^2
    + I(scale(age)^2)+I(scale(lwt)^2),trace=FALSE)
> print(birthwt.step2$anova)

Stepwise Model Path
Analysis of Deviance Table

Initial Model:
low ~ age + lwt + race + smoke + ptd + ht + ui + ftv

Final Model:
low ~ age + lwt + smoke + ptd + ht + ui + ftv
            + age:ftv + smoke:ui

        Step Df  Deviance Resid. Df Resid. Dev      AIC
1                              178   195.4755 217.4755
2   + age:ftv  2 12.474896     176   183.0006 209.0006
3 + smoke:ui  1  3.056805      175   179.9438 207.9438
4      - race  2  3.129586      177   183.0734 207.0734
```

Análisis de devianzas:

```
> print(anova(birthwt.step2))

Analysis of Deviance Table
Model: binomial, link: logit
Response: low
Terms added sequentially (first to last)

         Df Deviance Resid. Df Resid. Dev
NULL                       188     234.67
age       1   2.7600       187     231.91
lwt       1   4.7886       186     227.12
smoke     1   4.2440       185     222.88
ptd       1  10.6571       184     212.22
ht        1   6.8297       183     205.39
ui        1   2.1523       182     203.24
ftv       2   2.0135       180     201.23
age:ftv   2  14.2407       178     186.99
smoke:ui  1   3.9127       177     183.07
```

Note que aunque las variables `age` y `ftv` fueron descartadas previamente sus interacciones ahora se incluyen. Las pendientes de `age` difieren para los 3 niveles de `ftv`.

Problema de recién nacidos bajos de peso (cont.)

Modelo final:

Ajuste:

```
> print(summary(birthwt.step2))

 Call:
 glm(formula = low ~ age + lwt + smoke + ptd + ht + ui + ftv +
 age:ftv + smoke:ui, family = binomial, data = bwt)

 Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
 (Intercept)      -0.582374   1.421613  -0.410 0.682057
 age               0.075539   0.053967   1.400 0.161599
 lwt              -0.020373   0.007497  -2.717 0.006580
 smokeTRUE         0.780044   0.420385   1.856 0.063518
 ptdTRUE           1.560317   0.497001   3.139 0.001693
 htTRUE            2.065696   0.748743   2.759 0.005800
 uiTRUE            1.818530   0.667555   2.724 0.006446
 ftv1              2.921088   2.285774   1.278 0.201270
 ftv2+             9.244907   2.661497   3.474 0.000514
 age:ftv1         -0.161824   0.096819  -1.671 0.094642
 age:ftv2+        -0.411033   0.119144  -3.450 0.000561
 smokeTRUE:uiTRUE -1.916675   0.973097  -1.970 0.048877

 Null deviance: 234.67  on 188  degrees of freedom
 Residual deviance: 183.07  on 177  degrees of freedom
 AIC: 207.07

 Number of Fisher Scoring iterations: 5
```

```
> print(table(bwt$low,predict(birthwt.step2)>0))

    FALSE TRUE
  0   116   14
  1    28   31
```

–ebz

Problema de recién nacidos bajos de peso (cont.)

Comentario final:

El criterio AIC penaliza el número de términos menos severamente que como lo hace el estadístico de razón-de-verosimilitud o la prueba de Wald, por lo que aunque incluir la interacción `smoke:ui` reduce el AIC su correspondiente valor-*p* es apenas del 4.9 %. Se consideraron interacciones de tercer orden pero no fueron significativas.

Los residuales cuando la respuesta es una variable binaria no son informativas. De cualquier manera nos hubo grandes sorpresas en este análisis.

La capacidad de pronóstico del modelo se puede resumir con la siguiente tabla

peso	predicción	
observado	no bajo	bajo
no bajo	116	14
bajo	28	31

−ebz

Bibliografía

- De Jong, P. and Heller, D. Z. (2008) *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.
- Dobson, A. J. (1990) *An Introduction to Generalized Linear Models*. London: Chapman and Hall.
- Faraway, J. J. (2006) *Extending the Linear Model with R*. Boca Raton: Chapman and Hall.
- Hosmer, D. W. and Lemshow, S. (1989) *Applied Logistic Regression*. New York: Wiley.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. 2nd edition. London: Chapman and Hall.
- Montgomery, D. C., Peck E. A. and Vining, G. G. (2001) *Introduction to Lineal Regression Analysis*. 3rd edition. New York: Wiley.
- Myers, R. H., Montgomery D. C. and Vining, G. G. (2002) *Generalized Linear Models*. New York: Wiley.
- Pierce, D. A. and Schaffer, D. W. (1986) “Residuals in Generalized Linear Models”. *Journal of the American Statistical Association*. Vol. 81. No. 396. pp 977–986.
- Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. 4th edition. New York: Springer.

−ebz