Sesión 1: Especificaciones

Modelos Lineales Generalizados. Tema 1

Grado de Estadística

04/09/2018

Terminología

- Variable *explicativa* o *predictora*
- Variable *respuesta* o *dependiente*
- Modelos Lineales Generalizados *(Nelder i Wedderburn, 1972)*

Tipos de Variable

- **Categóricas Nominales**: Binarias o dicotómicas (2 categorías) y politómicas (más de 2 categorías). Son cualitativas.
- **Categóricas Ordinales**: Con ordenación natural entre las categorías. Pueden provenir de la discretización de una variable contínua. Son cualitativas.
- **Numéricas**: Pueden ser discretas (recuentos) o continuas. Son cuantitativas.
- **Factor**: variable explicativa cualitativa. Las distintas categorías se denominan *niveles*
- **Covariable**: variable explicativa cuantitativa.

Introducción a los Modelos Lineales Generalizados (MLGz)

Objetivo

Obtener un modelo matemático que permita reemplazr los datos $\mathbf{y}$ por los valores obtenidos del modelo $\hat{\boldsymbol{\mu}}$ en base a la optimización de un criterio estadístico: mínimos cuadrados, máxima verosimilitud, norma 1 (valor absoluto), etc. . .

Consideraciones

- Criterio de parsimonia
- **Model scope**: Rango de valores que facilitan buenas predicciones
- Validación de modelos: necesaria para garantizar la inferencia y predicción
- Valoración de modelos: criterios para comparar entre modelos validados

Tipos de Modelos

Distribución condicional	Dependencia Lineal	Dependencia No Lineal
Gaussiana	ML	MNL
$Y \| X \sim N_{\mu,\sigma^2}$	$Y = X\beta + \epsilon$	$Y = f(X, \beta) + \epsilon$
Familia exponencial	MLGz	MNLGz
$Y \| X \sim F_{\mu,\phi}$	$g(E(Y \| X)) = X\beta$	$g(E(Y \| X)) = f(X, \beta)$

Tipos de Modelos

Variables Explicativas	Variable de respuesta				
	Binaria	*Politómica*	*Cuantitativa Discreta*	*Cuantitativa Continua*	
				Normal	*Tiempo entre eventos*
Binaria	Tablas de contingencia Regresión logística Modelos log-lineales	Tablas de contingencia * Modelos log-lineales	Modelos log-lineales	Tests en medias de 2 grupos: t.test	Análisis de la Supervivencia
Politómicas	Tablas de contingencia Regresión logística Modelos log-lineales	Tablas de contingencia Modelos log-lineales	Modelos log-lineales	ONEWAY, ANOVA	Análisis de la Supervivencia
Continuas	Regresión logística	*	Modelos log-lineales	Regresión Múltiple	Análisis de la Supervivencia
Factores y covariables	Regresión logística	*	Modelos log-lineales	ANCOVA	Análisis de la Supervivencia
Efectos Aleatorios	Modelos mixtos	Modelos mixtos	Modelos mixtos	Modelos mixtos	Modelos mixtos

Estimación de Modelos Lineales Generalizados

- **Estimación**: criterio de máxima verosimilitud (permite incorporar la distribución asintótica de los estimadores)
- **Valoración**: bondad de ajuste evaluada con la devianza escalada (*scaled deviance*), discrepancia (*observed data vs fitted data*)

$$D'(\mathbf{y}, \boldsymbol{\mu}) = 2\ell(\mathbf{y}; \mathbf{y}) - 2\ell(\boldsymbol{\mu}; \mathbf{y})$$

donde:

$$\ell(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^{n} \log f(y_i, \mu_i) \qquad \boldsymbol{\mu}' = (\mu_1, ..., \mu_n) \quad \mathbf{y}' = (y_1, ..., y_n)$$

Componentes de los Modelos Lineales Generalizados

Son una extensión de los modelos lineales clásicos

Las observaciones $\mathbf{y}' = (y_1, ..., y_n)$ son realizaciones de un vector aleatorio $\mathbf{Y}' = (Y_1, ..., Y_n)$, cuyas componentes son estadísticamente independientes y distribuidas con medias $\boldsymbol{\mu}' = (\mu_1, ..., \mu_n)$

1. La **componente aleatoria** asume independencia entre las componentes de $\mathbf{Y}$ y pertenencia a distribuciones de la familia exponencial con valor esperado $E[\mathbf{Y}] = \boldsymbol{\mu}$

Componentes de los Modelos Lineales Generalizados (2)

2. La **componente sistemática** representada por el predictor lineal ($\boldsymbol{\eta}$) construido a partir de un número reducido de parámetros a estimar ($\boldsymbol{\beta}$) y de las variables regresoras ($\mathbf{X}$). Refleja la relación entre los predictores y el valor del predictor lineal.

   Matriz de diseño:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & .. & x_{1,p} \\ : & & : \\ x_{n,1} & .. & x_{n,p} \end{bmatrix}$$

   Coeficientes del modelo: $\boldsymbol{\beta}' = (\beta_1, ..., \beta_p)$

   Predictor Lineal: $\boldsymbol{\eta} = \mathbf{X}\beta$

Componentes de los Modelos Lineales Generalizados (3)

3. La **función de enlace** (*link function*): Relaciona el predictor lineal $\boldsymbol{\mu}$ con el valor esperado de la respuesta condicionada al volor que toman las variables predictoras.

$$g(E(\mathbf{Y}|\mathbf{X})) = \boldsymbol{\eta} = \mathbf{X}\beta$$

   La función de enlace es invertible y supone un mapeo biyectivo entre los valores del predictor lineal (que pueden tomar cualquier valor real) y el valor esperado de la distribución condicional de la respuesta (que suele tener restricciones)

Familia Exponencial de distribuciones

Sea Y una variable aleatoria cuya función de densidad (si es absolutamente continua) o función de masa de probabilidad (si es discreta) tiene la siguiente forma:

$$f_Y(y;\theta,\phi) = exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

donde *a(.),b(.),c(.)* son funciones específicas

- θ es el parámetro *natural* o *canónico* y determina el valor esperado de la distribución
- ϕ es el parámetro de *dispersión* y en algún caso no existe (o se asume que $\phi = 1$). Afecta a la varianza de la distribución
- La función de log-verosimilitud es:

$$\ell(\theta,\phi;y) = \log f_Y(y;\theta,\phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)$$

Familia Exponencial de distribuciones (2)

- La función *a(ϕ)* suele ser de la forma: $a(\phi) = \phi/w$ donde w es un *peso a priori*, habitualmente conocido
- La función *b(θ)* se denomina función *cumulante* y sus derivadas determinan, entre otros, la esperanza y varianza de la distribución:

  $$E(Y) = b'(\theta)$$

  $$V(Y) = a(\phi)b''(\theta) = a(\phi)v(\theta)$$

  donde $v(\theta)$ se denomina *función de varianza*.

La demostración se basa en dos propiedades de la familia exponencial:

- $E\left(\frac{\partial \ell(\theta)}{\partial \theta}\right) = 0$
- $E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right) + E\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^2 = 0$

Ejemplo 1: distribución Normal, $N(\mu, \sigma^2)$

$$
\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\
&= exp\left(\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + log(2\pi\sigma^2)\right)\right)
\end{aligned}
$$

y por lo tanto se puede deducir que:

$$\theta = \mu$$

- $a(\phi) = \sigma^2 = \phi$
- $b(\theta) = \frac{\mu^2}{2} = \frac{\theta^2}{2}$
- $c(y, \phi) = -\frac{1}{2}\left(\frac{y^2}{\phi} + log(2\pi\phi)\right)$

Ejemplo 1: distribución Normal, $N(\mu, \sigma^2)$

- $E(Y) = b'(\theta) = \theta = \mu$
- $V(Y) = a(\phi)b''(\theta) = \phi = \sigma^2$

Log-verosimilitud

$$\ell(\theta, \phi; y) = \frac{y\theta - \theta^2/2}{\phi} - \frac{1}{2}\left(\frac{y^2}{\phi} + log(2\pi\phi)\right)$$

$$\ell(\mu, \sigma^2; y) = \frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + log(2\pi\sigma^2)\right)$$

Deviancia escalada

$$D'(y, \mu) = \frac{(y - \mu)^2}{\sigma^2}$$

Ejemplo 2: distribución Poisson, *Pois*(λ)

$$f_Y(y) = e^{-\lambda}\frac{\lambda^y}{y!}$$
$$= exp(y\log\lambda - \lambda + \log(y!))$$

y por lo tanto se puede deducir que:

$$\theta = \log\lambda \Rightarrow \lambda = e^\theta$$

-$a(\phi) = 1$

-$b(\theta) = \lambda = e^\theta$

-$c(y,\phi) = \log(y!)$

Ejemplo 2: distribución Poisson, $Pois(\lambda)$

-$E(Y) = b'(\theta) = e^{\theta} = \lambda$

-$V(Y) = a(\phi)b''(\theta) = e^{\theta} = \lambda$

Log-verosimilitud

$$\ell(\theta; y) = y\theta - e^{\theta} - \log(y!)$$

$$\ell(\mu; y) = y\log(\mu) - \mu - \log(y!)$$

Deviancia escalada

$$D'(y, \mu) = 2\left(y\log\left(\frac{y}{\mu}\right) - (y - \mu)\right)$$

Ejemplo 3: distribución Binomial, $B(n, \pi)$ (con n conocida)

$$f_Y(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$$
$$= exp\left(y \log\left(\frac{\pi}{1-\pi}\right) - n \log\left(\frac{1}{1-\pi}\right) + \log\binom{n}{y} \right)$$

y por lo tanto se puede deducir que:

$$\theta = \log\left(\frac{\pi}{1-\pi}\right) \Rightarrow \pi = \frac{e^\theta}{1+e^\theta} = \frac{1}{1+e^{-\theta}}$$

- $a(\phi) = 1$
- $b(\theta) = n \log\left(\frac{1}{1-\pi}\right) = n \log(1+e^\theta)$
- $c(y, \phi) = \log\binom{n}{y}$

Ejemplo 3: distribución Binomial, $B(n, \pi)$ (con n conocida)

- $E(Y) = b'(\theta) = n\frac{e^\theta}{1+e^\theta} = n\pi$
- $V(Y) = n\frac{e^\theta}{(1+e^\theta)^2} = n\pi(1-\pi)$

Log-verosimilitud

$$\ell(\theta; y) = y\theta - n\log(1+e^\theta) + \log\binom{n}{y}$$

$$\ell(\mu; y) = y\log\left(\frac{\pi}{1-\pi}\right) - n\log\left(\frac{1}{1-\pi}\right) + \log\binom{n}{y} =$$

$$y\log\left(\frac{\mu}{n}\right) + (n-y)\log\left(\frac{n-\mu}{n}\right) + \binom{n}{y}$$

Deviancia escalada

$$D'(y, \mu) = 2\left(y\log\left(\frac{y}{\mu}\right) + (n-y)\log\left(\frac{n-y}{n-\mu}\right)\right)$$

Función de enlace (*link function*)

El parámetro "valor esperado" μ puede estar sujeto a restricciones:

- $X \sim Pois(\mu) \Rightarrow \mu \in [0, +\infty)$
- $X \sim B(n, \pi) \Rightarrow \pi \in [0, 1]$

El predictor lineal η puede tomar cualquier valor real.

La función de enlace relaciona de forma biyectiva el parámetro "valor esperado" con el predictor lineal:

$$g(\mu) = \eta \qquad \mu = g^{-1}(\eta)$$

Función de enlace canónico

Si la función de enlace transforma el parámetro "valor esperado" (μ) en el parámetro natural (θ), la función de enlace se denomina **enlace canónico**. El predictor lineal es un estimador del parámetro natural y los efectos sistemáticos tienen un efecto aditivo en la escala del parámetro natural

$$\theta = g(\mu) = \eta = X\beta$$

Como $\mu = b'(\theta)$, el enlace canónico corresponde a: $g(\mu) = (b')^{-1}(\mu)$ Ejemplos de enlaces canónicos:

- $X \sim N(\mu, \sigma^2) \Rightarrow g(\mu) = \mu$
- $X \sim Pois(\mu) \Rightarrow g(\mu) = \log(\mu)$
- $X \sim B(n, \pi) \Rightarrow g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$

Medidas de bondad de ajuste

Devianza escalada $D'(y, \mu)$: compara la función de log-verosimilitud del modelo a diagnosticar con modelo saturado (maximal) donde los parámetros esperados, $\hat{\boldsymbol{\mu}} = (\mu_1, ..., \mu_n)$ se sustituyen por las observaciones $\mathbf{y} = (y_1, ..., y_n)$:

$$D'(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2\ell(\mathbf{y}, \phi; \mathbf{y}) - 2\ell(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y})$$

Devianza: devianza escalada multiplicada por el parámetro de dispersión

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = D'(\mathbf{y}, \hat{\boldsymbol{\mu}})\phi$$

Ejemplos

- $X \sim N(\mu, \sigma^2) \Rightarrow D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$
- $X \sim Pois(\mu) \Rightarrow D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left(y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right)$
- $X \sim B(n, \pi)[\mu = n\pi] \Rightarrow D(\mathbf{y}, \hat{\boldsymbol{\mu}}) =$ $2\sum_{i=1}^{n}\left(y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (n - y_i)\log\left(\frac{n-y_i}{n-\hat{\mu}_i}\right)\right)$

Medidas de bondad de ajuste (2)

Estadístico de Pearson Generalizado:

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

- $X \sim N(\mu, \sigma^2) \Rightarrow \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}^2}$
- $X \sim Pois(\mu) \Rightarrow \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$
- $X \sim B(n, \pi)[\mu = n\pi] \Rightarrow \chi^2 = \sum_{i=1}^{n} \frac{(y_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i(1-\hat{\pi}_i)}$

Las distribuciones de la devianza escalada y del estadístico de Peaason son una χ^2 (exacta en el caso normal, asintótica en el resto de distribuciones)

Análisis de Residuos

- Residuos "crudos" (*raw residuals*): diferencia entre el valor observado y el esperado. Residuos en la escala de la respuesta.

$$r_i = y_i - \hat{\mu}_i$$

- Residuos de Pearson: residuos estandarizados por su desviación estándar

$$r_{P\,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

- Residuos de Devianza: contribución de la observación a la devianza

$$r_{D\,i} = sign(y_i - \hat{\mu}_i)\sqrt{d_i} \qquad d_i = 2(\ell(y_i; y_i) - \ell(\hat{\mu}_i; y_i))$$

Por tanto, $\chi^2 = \sum_{i=1}^{n} {r_P}_i^2$ y $D' = \sum_{i=1}^{n} {r_D}_i^2$