Modelos de conteo

Gabriel V. Montes-Rojas

Modelo de regresión de Poisson

Supongamos que la variable dependiente puede tomar cualquier valor entero no negativo $y = 0, 1, 2....$ Estos modelos se llaman modelos de conteo (count models), en los cuales se quiere explicar cuantas veces ocurre un determinado evento.

Ejemplo: y podría ser cantidad de visitas al médico en un mes:
$y = 0$ sin visitas
$y = 1$ una visita
$y = 2$ dos visitas
$y = 3$ tres visitas
$\vdots$
$y = 20$ veinte visitas
$y = ...$ no hay límite superior

¿Tiene sentido usar una regresión MCO?

Modelo de regresión de Poisson

- Consideremos la distribución de Poisson

$$f[y=j]=exp(-\lambda)\lambda^{j}/j!,\ j=0,1,2...$$

  donde $y!$ es el factorial de y, o sea $y\times(y-1)\times(y-2)\times ...\times 3\times 2$.
- Probar que si $y\sim Poisson(\lambda)$, entonces $E(y)=\lambda$ y $V(y)=\lambda$.
- Ahora consideremos un modelo condicional con $\lambda(\boldsymbol{x})=exp(\boldsymbol{x\beta})$. Esto implica un modelo loglineal: $log\lambda(\boldsymbol{x})=\boldsymbol{x\beta}$.
- Modelo de regresión de Poisson:

$$P[y=j|\boldsymbol{x}]=exp\left[-exp(\boldsymbol{x\beta})\right]\left[exp(\boldsymbol{x\beta})\right]^{j}/j!$$

- Un supuesto restrictivo del modelo de Poisson es que

$$Var(y|\boldsymbol{x})=E(y|\boldsymbol{x}).$$

Modelo de regresión de Poisson

- $\ell_i(\boldsymbol{\beta}) = \sum_{j=1}^{\infty} 1[y_i = j] log(P[y_i = j|x]) = y_i(\boldsymbol{x}_i\boldsymbol{\beta}) - exp(\boldsymbol{x}_i\boldsymbol{\beta})$
- $\boldsymbol{s}_i(\boldsymbol{\beta}) = y_i\boldsymbol{x}_i' - exp(\boldsymbol{x}_i\boldsymbol{\beta})\boldsymbol{x}_i'$
- Ejercicio: Probar que $E[\boldsymbol{s}_i(\boldsymbol{\beta})|\boldsymbol{x}_i] = 0$.
- $\boldsymbol{H}_i(\boldsymbol{\beta}) = -exp(\boldsymbol{x}_i\boldsymbol{\beta})\boldsymbol{x}_i'\boldsymbol{x}_i$
- Para interpretar los parámetros notemos que para una x contínua $\frac{\partial E[y|\boldsymbol{x}]}{\partial x_j} = exp(\boldsymbol{x}_i\boldsymbol{\beta})\beta_j$. Entonces $\beta_j = \frac{\partial E[y|\boldsymbol{x}]}{\partial x_j}\frac{1}{E[y|\boldsymbol{x}]} = \frac{\partial logE[y|\boldsymbol{x}]}{\partial x_j}$. Así β_j se puede interpretar como la semi-elasticidad de $E[y|\boldsymbol{x}]$ con respecto a x_j.
- Para dummies, notemos que $E[y|\boldsymbol{x}, d = 1] = exp(\beta_1 + \beta_2 x_2 + ... + \beta_d 1)$, y $E[y|\boldsymbol{x}, d = 1] = exp(\beta_1 + \beta_2 x_2 + ... + \beta_d 0)$. Entonces el efecto de la dummy es simplemente $exp(\beta_d)$.

Modelo de regresión de Poisson: STATA

- En STATA simplemente hay que escribir
  `poisson y x1 x2 x3`
- http://www.stata.com/manuals13/rpoisson.pdf
- http://www.stata.com/manuals13/rnbreg.pdf
- Ver
  http://www.ats.ucla.edu/stat/stata/examples/eacspd/chapter19.htm

Otros modelos de conteo (count data)

- El supuesto de media = varianza puede ser restrictivo. Entonces existen modelos alternativos que relajan esta condición.
- Binomial Negativa: asumamos que $f(y|\lambda) = exp(-\lambda)\lambda^y/y!$ pero $\lambda = \mu(x)\nu$ donde es una variable aleatoria. Notemos que $E(y|\mu) = \mu E(\nu)$ con lo cual si asumimos que $E(\nu) = 1$ estamos en el modelo binomial. Imponiendo distintas distribuciones sobre ν podemos lograr más flexibilidad.
- Asumamos que $\nu \sim Gamma$ (con un sólo parámetro), con densidad $g(\nu) = \nu^{\delta-1} exp(-\nu\delta)\delta^{\delta}/\Gamma(\delta)$, $\nu, \delta > 0$, $\Gamma(r) = \int_0^{\infty} z^{r-1} exp(-z)dz$, y con propiedades $E[\nu] = 1$, $V[\nu] = 1/\delta$. Si definimos $\alpha = 1/\delta$, entonces

$$E[y|\mu, \alpha] = \mu$$

$$V[y|\mu, \alpha] = \mu(1 + \alpha\mu)$$

- NegBin I asume $V[y|\mu, \alpha] = \mu(1 + \gamma)$, o sea, propocional a la media. NegBin II usa el modelo más general con una función cuadrática de la media.