

# Variables Enteras Positivas

Walter Sosa Escudero

wsosa@udesa.edu.ar

*Universidad de San Andrés*

## Introduccion

Características centrales:

- Variable explicada entera, positiva incluyendo el cero.  
Ejemplo: numero de hijos, visitas a un shopping en un periodo, visitas al medico, etc.
- No hay un valor maximo obvio o limite superior.
- Acumulacion en valores bajos.
- Muchos ceros.

*Ejemplo: publicaciones de estudiantes doctorales*

| Articles in  <br>last 3 yrs  <br>of PhD. | Freq. | Percent | Cum.   |
|--|-------|---------|--------|
| 0  | 275   | 30.05   | 30.05  |
| 1  | 246   | 26.89   | 56.94  |
| 2  | 178   | 19.45   | 76.39  |
| 3  | 84    | 9.18    | 85.57  |
| 4  | 67    | 7.32    | 92.90  |
| 5  | 27    | 2.95    | 95.85  |
| 6  | 17    | 1.86    | 97.70  |
| 7  | 12    | 1.31    | 99.02  |
| 8  | 1     | 0.11    | 99.13  |
| 9  | 2     | 0.22    | 99.34  |
| 10                                       | 1     | 0.11    | 99.45  |
| 11                                       | 1     | 0.11    | 99.56  |
| 12                                       | 2     | 0.22    | 99.78  |
| 16                                       | 1     | 0.11    | 99.89  |
| 19                                       | 1     | 0.11    | 100.00 |
| Total                                    | 915   | 100.00  |        |

El objetivo consiste en estimar un modelo adecuado para  $E(y|x)$  en donde  $x$  es un vector de variables explicativas.

Estudiaremos modelos de la siguiente forma:

$$E(y|x) = \exp(x'\beta)$$

- Garantizan que  $E(y|x) > 0$
- $\beta$  tiene una interpretacion simple ya que:

$$\frac{\partial \ln E(y|x)}{\partial x_k} = \beta_k$$

de modo que los coeficientes tienen interpretaciones de *semielasticidades*. Los coeficientes de variables binarias se interpretan como cambios proporcionales.

## La distribución de Poisson

$Y$  tiene *distribucion de Poisson* si:

$$f(y) = Pr(Y = y) = \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots$$

Es facil mostrar que:

$$E(Y) = V(Y) = \mu$$

La propiedad de que la esperanza coincide con la varianza se llama *equidispersion*.

## Regresion de Poisson

El modelo de *regresion de Poisson* corresponde a:

$$f(y/x) = \frac{e^{-\mu(x)} \mu(x)^y}{y!} \quad y = 0, 1, 2, \dots$$

$\mu(x) = \exp(x'\beta)$ ,  $x$  es un vector de  $K$  variables explicativas incluyendo un intercepto.

Consecuentemente:  $E(Y|x) = \exp(x'\beta)$

## Estimacion MV del modelo de Poisson

Supongamos  $y|x \sim Po(\mu = \exp(x'\beta_0))$ , y que disponemos de una muestra aleatoria i.i.d  $(y_i, x_i), i = 1, \dots, n$ . La funcion de verosimilitud sera:

$$L(\beta) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

con  $\mu_i \equiv \exp(x_i'\beta)$ . Tomando logaritmos:

$$l(\beta) = \sum_{i=1}^n \left[ y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i! \right]$$

El score para la muestra es:

$$s(\beta) = \sum_{i=1}^n s_i(\beta) = \sum_{i=1}^n \left[ y_i - \exp(x_i' \beta) \right] x_i$$

De modo que el estimador MV,  $\hat{\beta}_p$ , satisface:

$$\sum_{i=1}^n [y_i - \exp(x_i' \hat{\beta}_p)] x_i = 0$$

Propiedades:

- Consistencia:  $\hat{\beta}_p \xrightarrow{p} \beta_0$
- Normalidad asintótica:  $\sqrt{n}(\hat{\beta}_P - \beta_0) \xrightarrow{d} N(0, I(\beta_0)^{-1})$ ,  
 $I(\beta_0)$  es la matriz de información poblacional.
- Eficiencia asintótica.

En este caso:

$$\begin{aligned} I(\beta_0) &= E [s_i(\beta_0)s_i(\beta_0)'] \\ &= E [(y_i - \exp(x_i'\beta_0))^2 x_i x_i'] \\ &= E [(y - \exp(x'\beta_0))^2 x x'] \quad \text{Porque?} \end{aligned}$$

Bajo especificacion correcta:  $I(\beta_0) = -E(H(\beta_0))$

$$\begin{aligned} -E[H(\beta_0)] &= -E \left[ \frac{\partial}{\partial \beta} ([y_i - \exp(x_i'\beta)] x_i)_{\beta=\beta_0} \right] \\ &= E [\exp(x'\beta_0) x x'] \end{aligned}$$

Es facil chequear que:  $I(\beta_0) = -E(H(\beta_0))$  ya que bajo el supuesto de Poisson:

$$E [(y - \exp(x'\beta_0))^2 | x] = V(y|x) = E(y|x) = \exp(x'\beta_0)$$

Si  $\sqrt{n}(\hat{\beta}_P - \beta_0) \xrightarrow{d} N(0, I(\beta_0)^{-1})$ , entonces:

$$\hat{\beta}_p \sim N(\beta_0, [nI(\beta_0)]^{-1})$$

La varianza asintotica de  $\hat{\beta}_p$  es, entonces,  $[nI(\beta_0)]^{-1}$

Del resultado anterior,  $I(\beta_0)^{-1}$  puede ser estimada consistentemente usando el *metodo de matriz hessiana promedio*:

$$\frac{1}{n} \sum_{i=1}^n \exp(x'_i \hat{\beta}_P) x_i x'_i$$

lo que sugiere el siguiente estimador para la varianza asintotica:

$$\hat{V}_H = \left[ \sum_{i=1}^n \exp(x'_i \hat{\beta}_P) x_i x'_i \right]^{-1}$$

Tambien podemos usar el metodo del *producto externo del gradiente*. Un estimador consistente para  $I(\beta_0)$  sera:

$$\frac{1}{n} \sum_{i=1}^n \left( y_i - \exp(x'_i \hat{\beta}_P) \right)^2 x_i x'_i$$

de modo que un estimador valido para la varianza asintotica sera:

$$\hat{V}_{OP} = \left[ \sum_{i=1}^n \left( y_i - \exp(x'_i \hat{\beta}_P) \right)^2 x_i x'_i \right]^{-1}$$

## El estimador de Poisson como un estimador QMLE

QMLE = Cuasi maximo verosimil

- Que sucede con  $\hat{\beta}_p$  si la distribucion de los datos *no* es la de Poisson?
- Que sucede con  $\hat{V}_H$  y  $\hat{V}_{OP}$  si la distribucion de los datos *no* es la de Poisson?
- Mostraremos que si  $E(y|x) = \exp(x'\beta)$  sigue siendo valido,  $\hat{\beta}_P$  sigue siendo consistente y asintoticamente normal para *cualquier* distribucion subyacente.
- Mostraremos que  $\hat{V}_H$  y  $\hat{V}_{OP}$  no son consistentes, salvo condiciones muy particulares. Derivaremos nuevos estimadores consistentes.

- $\hat{\beta}_p$  definido implícitamente por:

$$\sum_{i=1}^n [y_i - \exp(x_i' \hat{\beta}_p)] x_i = 0$$

es un estimador *cuasi maximo-verosimil*: esta derivado de una funcion de verosimilitud no necesariamente cierta, pero asi y todo es consistente y asintoticamente normal.

- El ‘truco’ consiste en ver si a partir de esta definicion implicita es posible mostrar consistencia y normalidad asintotica *sin* recurrir a supuestos distributivos.

## Consistencia (bosquejo):

El argumento principal es similar al que usamos para MV:

- $\hat{\beta}_p$  es una raíz de  $\frac{1}{n}s(\beta)$ :  $\frac{1}{n} \sum_{i=1}^n [y_i - \exp(x'_i\beta)] x_i = 0$
- $\frac{1}{n}s(\beta) \xrightarrow{p} E[s_i(\beta)]$
- $E[s_i(\beta)] = 0$  solo si  $\beta = \beta_0$

Nuevamente, la idea es que  $\hat{\beta}_p$  es una raíz de una ecuación que tiene un límite en otra ecuación cuya raíz es  $\beta_0$ : si las ecuaciones convergen a un límite, las raíces convergen a la raíz del límite.

El primer punto es solo la definición del estimador, el segundo es una ley de grandes números.

El desafío es probar el tercero sin usar supuestos distributivos.

$$\begin{aligned} E[s_i(\beta)] &= E[(y_i - \exp(x'_i\beta))x_i] \\ &= E[E((y - \exp(x'\beta))x | x)] \\ &= E[(E(y|x) - \exp(x'\beta))x] = 0 \end{aligned}$$

si  $E(y|x) = \exp(x'\beta_0)$  y  $\beta = \beta_0$ .

Consecuentemente, la consistencia de  $\hat{\beta}_p$  se desprende de que  $E(y|x) = \exp(x'\beta_0)$  y no del supuesto de distribución de Poisson.

## Normalidad asintotica:

El argumento es identico al que hemos utilizado repetidas veces. Lo voy a dejar como ejercicio:

Resultado:

$$\sqrt{n}(\hat{\beta}_p - \hat{\beta}_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$

Con:  $A \equiv E(\nabla s_i(\beta_0))$ ,  $B \equiv E(s_i(\beta_0)s_i(\beta_0)')$ .

Si  $Y$  tiene distribucion de Poisson, notar que  $A = E(H(\beta_0))$  (la esperanza de la matriz Hessiana poblacional) y  $B = I(\theta_0)$  (la informacion poblacional), y que  $-A = B$ , de modo que, reemplazando,  $A^{-1}BA^{-1} = B^{-1}$ , que es el resultado que obtuvimos para el caso MV.

## Estimacion de la varianza asintotica de $\hat{\beta}_p$

$$\sqrt{n}(\hat{\beta}_p - \beta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$

$$\hat{\beta}_p \sim N\left(\beta_0, \frac{1}{n}A^{-1}BA^{-1}\right)$$

Notar que si:

$$s_i(\beta) = [y_i - \exp(x'_i\beta)]x_i$$

entonces:

$$\nabla s_i(\beta) = -\exp(x'_i\beta)x_ix'_i$$

Consecuentemente:

$$A \equiv E(\nabla s_i(\beta_0)) = E[-\exp(x'_i\beta_0)x_ix'_i] = -E[E(y|x)xx']$$

y

$$B \equiv E(s_i(\beta_0)s_i(\beta_0)') = E[(y - \exp(x'\beta_0))^2xx'] = E[V(y|x)xx']$$

,usando la ley de esperanzas iteradas (verificar)

Es facil verificar que:

$$\hat{A} = -\frac{1}{n} \sum_{i=1}^n \exp(x'_i\hat{\beta}_p)x_ix'_i \xrightarrow{p} A$$

y que

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n [y_i - \exp(x'_i\hat{\beta}_p)]^2 x_ix'_i \xrightarrow{p} B$$

lo deajo como ejercicio...

de modo que, reemplazando y simplificando:

$$\begin{aligned}\hat{V}_R &= \frac{1}{n} A^{-1} \hat{B} \hat{A}^{-1} \\ &= \left( \sum_{i=1}^n \hat{\mu}_i x_i x_i' \right)^{-1} \left( \sum_{i=1}^n \hat{\omega}_i^2 x_i x_i' \right) \left( \sum_{i=1}^n \hat{\mu}_i x_i x_i' \right)^{-1}\end{aligned}$$

,  $\hat{\mu}_i \equiv \exp(x_i' \hat{\beta}_p)$  y  $\hat{\omega}_i \equiv y_i - \exp(x_i' \hat{\beta}_p)$ , es un estimador valido para la varianza asintotica de  $\hat{\beta}_p$ .

## Algunas simplificaciones surgen de casos especiales:

- Si el supuesto de Poisson es verdadero, entonces  $A^{-1}BA^{-1} = -A^{-1}$ , consecuentemente, en este caso  $\hat{V}_H$  y  $\hat{V}_{OP}$  son consistentes. En caso contrario, no lo son.
- Si  $V(y|x) = E(y|x)$ , también  $A^{-1}BA^{-1} = -A^{-1}$ , entonces  $\hat{V}_H$  y  $\hat{V}_{OP}$  son consistentes. La consistencia depende del supuesto de equidispersión, no del de Poisson!
- Si  $V(y|x) = \phi E(y|x)$  (varianza proporcional a la esperanza), entonces,  $B = -\phi A$ , entonces:

$$A^{-1}BA^{-1} = \phi(-A)^{-1}$$

Notar que:

$$\phi = \frac{V(y|x)}{E(y|x)}$$

Entonces, un estimador consistente para  $\phi$  puede ser:

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

con  $\hat{\mu}_i \equiv \exp(x_i' \hat{\beta}_p)$  (verificar consistencia)

Consecuentemente, cuando  $V(y|x) = \phi E(y|x)$ , un estimador consistente para  $V(\beta_p)$  es:

$$\hat{V}_{GLM} = \hat{\phi} \left( -\hat{A} \right)^{-1} = \hat{\phi} \left( \sum_{i=1}^n \mu_i x_i x_i' \right)^{-1}$$

Notar que:

$$\hat{V}_{GLM} = \hat{\phi} \hat{V}_H(\hat{\beta}_p)$$

entonces, cuando  $\phi > 1$  (sobredispersión), el estimador MV para la varianza tiende a subestimar las verdaderas varianzas.

En síntesis...

- $\hat{\beta}_p$  es consistente y asintóticamente normal si  $E(y|x) = \exp(x'\beta_0)$ , para cualquier distribución.
- Bajo el supuesto de que la distribución es la de Poisson, también es asintóticamente eficiente (Porque?).
- $V_H$  y  $V_{OP}$  son consistentes si  $V(y|x) = E(y|x)$  (equidispersión), lo cual incluye como caso particular al de Poisson.
- $V_R$  es siempre consistente.
- Si  $V(y|x) = \phi E(y|x)$ , y  $\hat{\phi} \xrightarrow{p} \phi$ ,  $V_{GLM} = \hat{\phi} V_H$  es consistente.
- Bajo *sobredispersión* ( $\phi > 1$ ),  $V_H$  o  $V_{OP}$  subestiman la verdadera varianza (tienden a sugerir más precisión que la real).