Estadística Aplicada III

Modelos lineales Generalizados
Otros Temas

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 15: 27 de noviembre de 2018

Sobredispersión I

- Hemos visto que en un GLM en donde el parámetro de dispersión ϕ es constante (`binomial` o `poisson`), la devianza residual es:

$$Dev_{res} = 2(l_s - l_M) \dot{\sim} \chi^2_{(n-k)}$$

  donde l_s es la log-verosimilitud del modelo saturado, y l_M la del modelo ajustado (que tiene k coeficientes).
- Para una distribución $\chi^2_{(p)}$, sabemos que $E(\chi^2_{(p)}) = p$. Así que si un modelo ajusta bien los datos, se esperaría que $Dev_{res} \approx gl_{res}$.
- En el caso de que $Dev_{res} > gl_{res}$ esto se puede deber a dos principales causas:
  1. Un modelo mal ajustado:
     - El modelo no está bien especificado.
     - Hay algunos factores explicativos que no están incorporados en los predictores.
     - posiblemente outliers.
  2. Por **sobredispersión**: la variación observada de los datos es mayor que la predicha por el modelo.
     - La variabilidad de los casos a nivel individual es importante.
     - Hay correlación entre las respuestas.

Sobredispersión II

- El diseño muestral considera conglomerados.
- Se omiten algunas variables no observadas.
- Exceso de ceros.

Consecuencias de la sobredispersión I

- Usualmente se obtienen estimadores consistentes de β, pero
  - Los errores estándar no son correctos
  - Se obtienen intervalos de confianza demasiado optimistas.
  - Se pueden seleccionar modelos demasiado complejos.
- Las consecuencias pueden ser potencialmente severas.

Ejemplo: Modelo de vínculos (Ornstein) I

- El modelo de vínculos con empresas que vimos antes. En este ejemplo se tiene evidencia de sobredispersión, ya que la devianza residual es 1887 con 234 grados de libertad.

Ejemplo: Modelo de vínculos (Ornstein) II

```
data("Ornstein")
mod1 <- glm(interlocks ~ ., family = poisson, data = Ornstein)
summary(mod1)
```

```
Call:
glm(formula = interlocks ~ ., family = poisson, data = Ornstein)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.9908  -2.4767  -0.8582   1.3472   7.3610

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.325e+00  5.193e-02  44.762  < 2e-16 ***
assets       2.085e-05  1.202e-06  17.340  < 2e-16 ***
sectorBNK   -4.092e-01  1.560e-01  -2.623  0.00872 **
sectorCON   -6.196e-01  2.120e-01  -2.923  0.00347 **
sectorFIN    6.770e-01  6.879e-02   9.841  < 2e-16 ***
sectorHLD    2.085e-01  1.189e-01   1.754  0.07948 .
sectorMAN    5.260e-02  7.553e-02   0.696  0.48621
sectorMER    1.777e-01  8.654e-02   2.053  0.04006 *
sectorMIN    6.211e-01  6.690e-02   9.283  < 2e-16 ***
sectorTRN    6.778e-01  7.483e-02   9.059  < 2e-16 ***
sectorWOD    7.116e-01  7.532e-02   9.447  < 2e-16 ***
nationOTH   -1.632e-01  7.362e-02  -2.217  0.02663 *
nationUK    -5.771e-01  8.903e-02  -6.482 9.05e-11 ***
nationUS    -8.259e-01  4.897e-02 -16.867  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3737.0  on 247  degrees of freedom
Residual deviance: 1887.4  on 234  degrees of freedom
AIC: 2813.4

Number of Fisher Scoring iterations: 5
```

Modelos de quasi-verosimilitud: modelo quasiPoisson

- En los modelos lineales, cómo son los estimadores cuando aplican mínimos cuadrados a un modelo que no cumple el supuesto de normalidad?

Modelos de quasi-verosimilitud: modelo quasiPoisson

- En los modelos lineales, cómo son los estimadores cuando aplican mínimos cuadrados a un modelo que no cumple el supuesto de normalidad?
- los estimadores son insesgados, asintóticamente normales, y tienen la matriz de covarianzas usual, siempre que se cumplan los supuestos de linealidad, varianza del error constante e independencia.

Modelos de quasi-verosimilitud: modelo quasiPoisson

- En los modelos lineales, cómo son los estimadores cuando aplican mínimos cuadrados a un modelo que no cumple el supuesto de normalidad?
- los estimadores son insesgados, asintóticamente normales, y tienen la matriz de covarianzas usual, siempre que se cumplan los supuestos de linealidad, varianza del error constante e independencia.
- De hecho, los estimadores de mínimos cuadrados no son los de máxima verosimilitud, pero son maximales eficientes entre los estimadores lineales insesgados (por Gauss-Markov).

Modelos de quasi-verosimilitud: modelo quasiPoisson

- En los modelos lineales, cómo son los estimadores cuando aplican mínimos cuadrados a un modelo que no cumple el supuesto de normalidad?
- los estimadores son insesgados, asintóticamente normales, y tienen la matriz de covarianzas usual, siempre que se cumplan los supuestos de linealidad, varianza del error constante e independencia.
- De hecho, los estimadores de mínimos cuadrados no son los de máxima verosimilitud, pero son maximales eficientes entre los estimadores lineales insesgados (por Gauss-Markov).
- La idea esencial de la estimación quasi-verosímil es utilizar la misma función a maximizar para cualquier distribución que corresponda a la media y varianza (dos primeros momentos) de una distribución de una familia exponencial.

Modelos de quasi-verosimilitud: modelo quasiPoisson

- En los modelos lineales, cómo son los estimadores cuando aplican mínimos cuadrados a un modelo que no cumple el supuesto de normalidad?
- los estimadores son insesgados, asintóticamente normales, y tienen la matriz de covarianzas usual, siempre que se cumplan los supuestos de linealidad, varianza del error constante e independencia.
- De hecho, los estimadores de mínimos cuadrados no son los de máxima verosimilitud, pero son maximales eficientes entre los estimadores lineales insesgados (por Gauss-Markov).
- La idea esencial de la estimación quasi-verosímil es utilizar la misma función a maximizar para cualquier distribución que corresponda a la media y varianza (dos primeros momentos) de una distribución de una familia exponencial.
- McCullagh y Nelder demuestran que las propiedades de los estimadores de máxima verosimilitud para familias exponenciales se comparten con distribuciones que coinciden con los dos primeros momentos de una familia exponencial.

Modelos de quasi-verosimilitud: modelo quasiPoisson

- El modelo quasiPoisson introduce un parámetro de dispersión en la varianza: $Var(Y_i|\eta_i) = \phi\mu_i$.
- Si $\phi \neq 1$ la varianza condicional cambia más rápido (si $\phi > 1$) o lento (si $\phi < 1$) que la media.
- Los estimados de quasi-verosimilitud son idénticos a los de máxima verosimilitud, pero los errores estándar de los coeficientes cambian: si $\tilde{\phi}$ es el parámetro estimado para el modelo, entonces:

$$se(\hat{\beta}_i^{qP}) = \tilde{\phi}^{1/2} se(\hat{\beta}_i^{P})$$

- El estimador que se usa para el parámetro de dispersión es el de momentos que ya se mencionó antes:

$$\tilde{\phi} = \frac{1}{n-k}\sum \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

Ejemplo: Modelo quasiPoisson para los datos de vínculos I

- Repetimos el ejercicio cambiando la familia de `poisson` a `quasipoisson`.
- Esto no cambia los valores del ajuste, pero notemos que los errores ahora ya no son z sino t, lo que cambia la significancia de los coeficientes.
- También estima el parámetro de dispersión que ya no es 1, sino $\tilde{\phi} = 7.9439$ y entonces cada error estándar se multiplica por $\sqrt{7.9439} = 2.8184925$.

Ejemplo: Modelo quasiPoisson para los datos de vínculos II

```
modqP <- glm(interlocks ~ ., family = quasipoisson, data = Ornstein)
summary(modqP)


Call:
glm(formula = interlocks ~ ., family = quasipoisson, data = Ornstein)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.9908  -2.4767  -0.8582   1.3472   7.3610

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.325e+00  1.464e-01  15.881  < 2e-16 ***
assets       2.085e-05  3.389e-06   6.152 3.28e-09 ***
sectorBNK   -4.092e-01  4.397e-01  -0.931 0.353003
sectorCON   -6.196e-01  5.974e-01  -1.037 0.300779
sectorFIN    6.770e-01  1.939e-01   3.491 0.000574 ***
sectorHLD    2.085e-01  3.350e-01   0.622 0.534410
sectorMAN    5.260e-02  2.129e-01   0.247 0.805075
sectorMER    1.777e-01  2.439e-01   0.728 0.467056
sectorMIN    6.211e-01  1.886e-01   3.294 0.001142 **
sectorTRN    6.778e-01  2.109e-01   3.214 0.001493 **
sectorWOD    7.116e-01  2.123e-01   3.352 0.000936 ***
nationOTH   -1.632e-01  2.075e-01  -0.787 0.432335
nationUK    -5.771e-01  2.509e-01  -2.300 0.022339 *
nationUS    -8.259e-01  1.380e-01  -5.984 8.10e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 7.943873)

    Null deviance: 3737.0  on 247  degrees of freedom
Residual deviance: 1887.4  on 234  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Otro enfoque para sobredispersión: regresión binomial negativa I

- Recordemos lo siguiente: $Z \sim \mathcal{G}(\alpha, \beta)$ si tiene como densidad:

$$f(z) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} z^{\alpha-1} e^{-z/\beta}$$

  donde α es el parámetro de escala y β el de forma. Con esta parametrización, $\mathrm{E}(Z) = \alpha\beta$ y $\mathrm{Var}(Z) = \alpha\beta^2$.
- Otra manera de pensar el problema de sobredispersión es permitir que el parámetro de la distribución Poisson tenga su propio comportamiento. Usualmente se considera el siguiente modelo jerárquico:

$$\begin{aligned} Y|\lambda &\sim \mathcal{P}(\lambda) \\ \lambda &\sim \mathcal{G}(\omega, \mu_i/\omega) \end{aligned}$$

  de tal forma que $\mathrm{E}(\lambda) = \mu_i = \omega\mu_i/\omega$.

Otro enfoque para sobredispersión: regresión binomial negativa II

- A partir de este modelo jerárquico, ¿cuál es la distribución de Y_i?

$$P(Y = y) = \int_0^\infty f_Y(y|\lambda) f_\lambda(\lambda)\, d\lambda \quad = \quad \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \frac{\omega^\omega \lambda^{\omega-1} e^{-\lambda\omega/\mu_i}}{\mu_i^\omega \Gamma(\omega)}\, d\lambda$$
$$\ldots$$

- Entonces Y sigue la generalización de una distribución binomial negativa, que es la distribución de Polya:

$$P(Y = y_i) = \frac{\Gamma(y_i + \omega)}{y_i! \Gamma(\omega)} \left(\frac{\omega}{\omega + \mu_i}\right)^\omega \left(\frac{\mu_i}{\omega + \mu_i}\right)^{y_i}$$

  que tiene media $\mathrm{E}(Y_i) = \mu_i$ y varianza $\mathrm{Var}(Y_i) = \mu_i + \mu_i^2/\omega$.

- En el contexto de los modelos GLM's, el parámetro ω de la binomial negativa se supone conocido. Se puede construir un grid de valores para estimar aquel valor de ω que minimice el AIC.

Ejemplo: datos de vínculos con Binomial Negativa I

```
library(MASS)
modbn <- glm(interlocks ~ ., family = negative.binomial(1), data = Ornstein)
theta <- seq(0.5, 2.5, by=0.5) # grid
aics <- rep(0,5)
for (i in seq(along=theta)) aics[i] <- AIC(update(modbn, family=negative.binomial(theta[i])))
rbind(theta,aics)
```

```
          [,1]     [,2]     [,3]     [,4]     [,5]
theta    0.500    1.000    1.500    2.000    2.500
aics  1789.075 1721.215 1716.612 1730.192 1750.512
```

Entonces el mínimo AIC se tiene alrededor de $\omega = 1.5$, que nos da la estimación siguiente:

Ejemplo: datos de vínculos con Binomial Negativa II

```
modbnopt <- update(modbn,family=negative.binomial(1.5))
summary(modbnopt)



Call:
glm(formula = interlocks ~ ., family = negative.binomial(1.5),
    data = Ornstein)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7990  -1.1440  -0.2933   0.4637   2.1526

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.256e+00  1.425e-01  15.828  < 2e-16 ***
assets       3.252e-05  5.464e-06   5.950 9.69e-09 ***
sectorBNK   -1.057e+00  5.467e-01  -1.934  0.05437 .
sectorCON   -7.297e-01  4.580e-01  -1.593  0.11248
sectorFIN    6.094e-01  2.350e-01   2.593  0.01011 *
sectorHLD    1.397e-01  3.613e-01   0.387  0.69928
sectorMAN    7.790e-02  1.919e-01   0.406  0.68515
sectorMER    2.051e-01  2.407e-01   0.852  0.39498
sectorMIN    5.217e-01  1.891e-01   2.758  0.00627 **
sectorTRN    5.957e-01  2.470e-01   2.412  0.01666 *
sectorWOD    6.537e-01  2.401e-01   2.722  0.00697 **
nationOTH    8.405e-03  2.401e-01   0.035  0.97211
nationUK    -4.791e-01  2.447e-01  -1.958  0.05140 .
nationUS    -7.862e-01  1.367e-01  -5.751 2.77e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.5) family taken to be 1.02452)

    Null deviance: 487.78  on 247  degrees of freedom
Residual deviance: 322.28  on 234  degrees of freedom
AIC: 1716.6

Number of Fisher Scoring iterations: 9
```

Residuales y gráficas de residuales I

- En los GLMs no hay residuales como en el sentido típico de los modelos lineales. En los modelos lineales, los residuales son: $\hat{y} - y$ para representar el error estadístico $\epsilon = \mathrm{E}(y|\eta) - y$. Pero en los GLM's no hay componente aditivo.
- Hay varios tipos de residuales disponibles para GLMs:
  1. Residuales respuesta: $y_i - \hat{\mu}_i$. Este tipo de residuales son los usuales en el modelo gaussiano, pero no se pueden usar para diagnósticos, ya que ignoran que la varianza no es constante.
  2. Residuales Pearson: estos se basan en la prueba de bondad de ajuste del modelo. Son los que usualmente se usan con un GLM por su analogía directa con los modelos lineales:

$$e_{P,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var(y_i}|\mathbf{x})/\hat{\phi}}}$$

     Se obtienen en R con `residuals(modelo,type="pearson")`

Residuales y gráficas de residuales II

3. Residuales Pearson estandarizados: estos corrigen a los anteriores por la varianza condicional y por el leverage de las observaciones:

$$e_{PS,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{Var(y_i|\mathbf{x})}(1-h_i)}}$$

Los valores de h_i se toman de la última iteración del método IWLS. A diferencia de los modelos lineales, estos valores dependenden de y y de la configuración de los predictores.

4. residuales de la devianza, $e_{D,i}$: son las raíces cuadradas de los componentes caso por caso de la devianza residual, poniéndoles el signo de $y_i - \hat{\mu}_i$. Se obtienen de `R` con `residuals(modelo, type="deviance")`.

5. Residuales de la devianza estandarizados:

$$e_{DS,i} = \frac{e_{D,i}}{sqrt\hat{\phi}(1-h_i)}$$

6. Residuales estudentizados (aproximados):

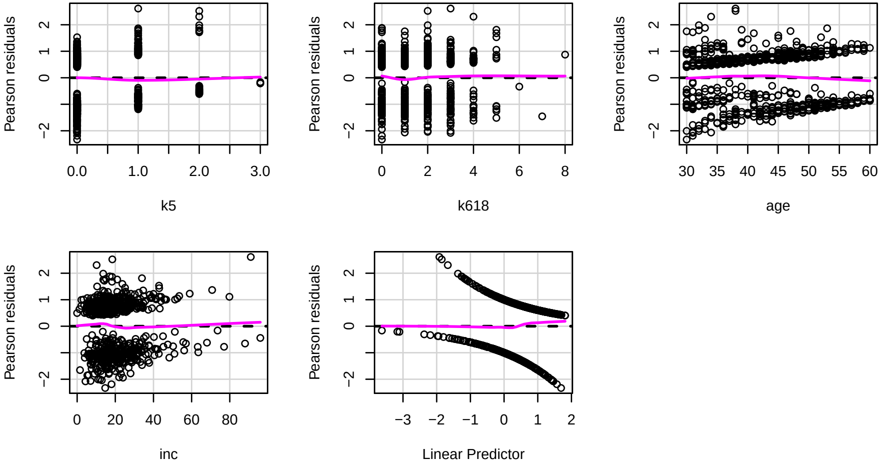
$$e_{T,i} = signo(y_i - \hat{\mu}_i)\sqrt{(1-h_i)e^2_{DS,i} + h_i e^2_{PS,i}}$$

Ejemplo: Fuerza laboral de las mujeres I

Por ejemplo, usando los datos de Mroz sobre fuerza laboral de las mujeres (que vimos en laboratorio la semana pasada)

```
library(car)
data(Mroz)
mod1 <- glm(lfp ~ k5 + k618 + age + inc, family = binomial(link=logit), data = Mroz)
residualPlots(mod1,layout=c(2,3))
```

Ejemplo: Fuerza laboral de las mujeres II



Ejemplo: Fuerza laboral de las mujeres III

```
     Test stat Pr(>|Test stat|)
k5      0.6868           0.4073
k618    0.2604           0.6098
age     1.2440           0.2647
inc     1.0664           0.3018
```