Modelos Lineales

Modelos Lineales Generalizados

Fernando Massa; Bruno Bellagamba

25 de junio 2024

FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN
IESTA INSTITUTO DE ESTADÍSTICA
UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Temario

- Definiremos la *devianza* como medida de bondad de ajuste.
- Veremos como se estiman los parámetros de un GLM no normal.
- Evaluaremos las predicciones de estos modelos, haciendo hincapié en la regresión logística.

En los modelos lineales vistos anteriormente, medíamos la *bondad de ajuste* del modelo a través de varios indicadores, siendo el R^2 el que nos brindaba una interpretación más clara.

En los modelos lineales vistos anteriormente, medíamos la *bondad de ajuste* del modelo a través de varios indicadores, siendo el R^2 el que nos brindaba una interpretación más clara.

No obstante, si bien existen versiones de este indicador para casos de *GLM* donde Y no es normal, estos tienen algunos inconvenientes que dificultan su uso.

En los modelos lineales vistos anteriormente, medíamos la *bondad de ajuste* del modelo a través de varios indicadores, siendo el R^2 el que nos brindaba una interpretación más clara.

No obstante, si bien existen versiones de este indicador para casos de GLM donde Y no es normal, estos tienen algunos inconvenientes que dificultan su uso.

Ya que los GLM son una generalización del modelo lineal general, veremos que es posible obtener una medida de bondad de ajuste que generaliza el R^2,

En los modelos lineales vistos anteriormente, medíamos la *bondad de ajuste* del modelo a través de varios indicadores, siendo el R^2 el que nos brindaba una interpretación más clara.

No obstante, si bien existen versiones de este indicador para casos de *GLM* donde Y no es normal, estos tienen algunos inconvenientes que dificultan su uso.

Ya que los *GLM* son una generalización del modelo lineal general, veremos que es posible obtener una medida de bondad de ajuste que generaliza el R^2, esta medida es la **Devianza**.

Devianza

Intuitivamente decimos que la *devianza* mide que tan bien se ajusta un cierto *GLM* a los datos, con respecto a un modelo ideal que tiene un ajuste *perfecto* (que tiene tantos parámetros como observaciones). A este modelo ideal se lo suele conocer como modelo saturado. De esta manera:

- La devianza es una medida de error: valores bajos corresponden a situaciones donde el ajuste del modelo está cerca del modelo ideal.
- Análogamente, valores altos de la devianza indican un peor ajuste.

Devianza

Intuitivamente decimos que la *devianza* mide que tan bien se ajusta un cierto *GLM* a los datos, con respecto a un modelo ideal que tiene un ajuste *perfecto* (que tiene tantos parámetros como observaciones). A este modelo ideal se lo suele conocer como modelo saturado. De esta manera:

- La devianza es una medida de error: valores bajos corresponden a situaciones donde el ajuste del modelo está cerca del modelo ideal.
- Análogamente, valores altos de la devianza indican un peor ajuste.

Formalmente diremos que la devianza se plantea:

$$D\left(Y, \hat{\mu}\right) = 2 \times \left[\mathcal{L}(Y|Y, \phi) - \mathcal{L}(Y|\hat{\mu}, \phi)\right]$$

Devianza

Intuitivamente decimos que la *devianza* mide que tan bien se ajusta un cierto *GLM* a los datos, con respecto a un modelo ideal que tiene un ajuste *perfecto* (que tiene tantos parámetros como observaciones). A este modelo ideal se lo suele conocer como modelo saturado. De esta manera:

- La devianza es una medida de error: valores bajos corresponden a situaciones donde el ajuste del modelo está cerca del modelo ideal.
- Análogamente, valores altos de la devianza indican un peor ajuste.

Formalmente diremos que la devianza se plantea:

$$D(Y, \hat{\mu}) = 2 \times [\mathcal{L}(Y|Y,\phi) - \mathcal{L}(Y|\hat{\mu},\phi)]$$

Siendo $\mathcal{L}(Y|\hat{\mu},\phi)$ la log-verosimilitud evaluada en $\hat{\mu}$ y $\mathcal{L}(Y|Y,\phi)$ la log-verosimilitud del modelo *saturado*.

Devianza

Alternativamente, la devianza se puede plantear como:

$$D\left(Y, \hat{\mu}(\mathbf{x})\right) = -2 \times \log\left[\frac{L(Y|\hat{\mu}, \phi)}{L(Y|Y, \phi)}\right]$$

Siendo $L(Y|\hat{\mu}, \phi)$ y $L(Y|Y\phi)$ la verosimilitud de los modelos propuesto y saturado respectivamente.

Devianza

Alternativamente, la devianza se puede plantear como:

$$D\left(Y, \hat{\mu}(\mathbf{x})\right) = -2 \times \log\left[\frac{L(Y|\hat{\mu}, \phi)}{L(Y|Y, \phi)}\right]$$

Siendo $L(Y|\hat{\mu}, \phi)$ y $L(Y|Y\phi)$ la verosimilitud de los modelos propuesto y saturado respectivamente.

El término entre paréntesis es el cociente entre la verosimilitud del modelo y la máxima verosimilitud posible.

Devianza

Alternativamente, la devianza se puede plantear como:

$$D\left(Y, \hat{\mu}(\mathbf{x})\right) = -2 \times \log \left[\frac{L(Y|\hat{\mu}, \phi)}{L(Y|Y, \phi)} \right]$$

Siendo $L(Y|\hat{\mu}, \phi)$ y $L(Y|Y\phi)$ la verosimilitud de los modelos propuesto y saturado respectivamente.

El término entre paréntesis es el cociente entre la verosimilitud del modelo y la máxima verosimilitud posible. Es la proporción de verosimilitud explicada por el modelo.

Alternativamente, la devianza se puede plantear como:

$$D\left(Y, \hat{\mu}(\mathbf{x})\right) = -2 \times \log\left[\frac{L(Y|\hat{\mu},\phi)}{L(Y|Y,\phi)}\right]$$

Siendo $L(Y|\hat{\mu},\phi)$ y $L(Y|Y\phi)$ la verosimilitud de los modelos propuesto y saturado respectivamente.

El término entre paréntesis es el cociente entre la verosimilitud del modelo y la máxima verosimilitud posible. Es la proporción de verosimilitud explicada por el modelo.

Existen varias propuestas de *pseudo*-R^2 para los *GLM*. Las más conocidas son:

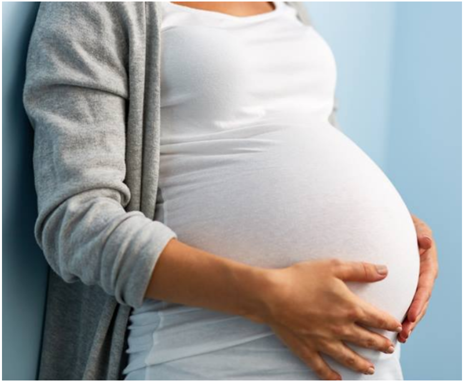
- **McFadden**: $1 - \frac{\mathcal{L}(\beta)}{\mathcal{L}(\beta_0)}$
- **Cox & Snell**: $1 - \left(\frac{L(Y|\beta_0,\phi)}{L(Y|\beta,\phi)}\right)^{2/n}$
- **Nagelkerke**: $\frac{1-\left[\frac{L(Y|\beta_0,\phi)}{L(Y|\beta,\phi)}\right]^{2/n}}{1-L(Y|\beta_0,\phi)^{2/n}}$

Ejemplo

Veamos el ejemplo donde se busca modelar la probabilidad (o la chance) de que un niño tenga bajo peso al nacer en función de ciertas características y conductas de la madre.

Ejemplo

Veamos el ejemplo donde se busca modelar la probabilidad (o la chance) de que un niño tenga bajo peso al nacer en función de ciertas características y conductas de la madre.



Vayamos al R y veamos como obtener estos indicadores.

En el curso de inferencia I definimos la función *score* $s(\theta)$ como:

$$\begin{aligned} \ell(y|\theta,\phi) &= \log f_Y(y|\theta,\phi) \\ s(\theta) &= \tfrac{\partial \ell(y|\theta,\phi)}{\partial \theta} \end{aligned}$$

Si bien esta es la definición para cualquier función de densidad, nos interesa el caso en el que $f_Y(y|\theta,\phi)$ es una densidad perteneciente a la familia exponencial, es decir:

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

PROPIEDADES DEL *score*

En el curso de inferencia I definimos la función *score* $s(\theta)$ como:

$$\begin{aligned} \ell(y|\theta,\phi) &= \log f_Y(y|\theta,\phi) \\ s(\theta) &= \frac{\partial \ell(y|\theta,\phi)}{\partial \theta} \end{aligned}$$

Si bien esta es la definición para cualquier función de densidad, nos interesa el caso en el que $f_Y(y|\theta,\phi)$ es una densidad perteneciente a la familia exponencial, es decir:

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

Sea quien sea $f_y(y)$ es posible probar (y queda como ejercicio para entregar) que:

$$\begin{aligned} \mathbb{E}\left(\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right) &= 0 \\ \mathbb{E}\left(\frac{\partial^2 \ell(y|\theta,\phi)}{\partial \theta^2}\right) + \mathbb{E}\left[\left(\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right)^2\right] &= 0 \end{aligned}$$

Propiedades de la función $b(\theta)$

Ahora si, asumiendo que $f_Y(y|\theta, \phi)$ pertenece a la familia exponencial:

Propiedades de la función $b(\theta)$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right] = 0$$

Propiedades de la función $b(\theta)$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right] &= 0 \\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial \theta}\right] &= 0
\end{aligned}$$

Ahora si, asumiendo que $f_Y(y|\theta, \phi)$ pertenece a la familia exponencial:

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right] &= 0 \\ \mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial \theta}\right] &= 0 \\ \mathbb{E}\left\{\frac{\partial}{\partial \theta}\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]\right\} &= 0 \end{aligned}$$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial \theta}\right] &= 0 \\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial \theta}\right] &= 0 \\
\mathbb{E}\left\{\frac{\partial}{\partial \theta}\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]\right\} &= 0 \\
\mathbb{E}\left[\frac{y - b^{'}(\theta)}{a(\phi)}\right] &= 0
\end{aligned}
$$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial\theta}\right] &= 0\\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial\theta}\right] &= 0\\
\mathbb{E}\left\{\frac{\partial}{\partial\theta}\left[\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right]\right\} &= 0\\
\mathbb{E}\left[\frac{y-b^{'}(\theta)}{a(\phi)}\right] &= 0\\
\frac{\mathbb{E}(y)-b^{'}(\theta)}{a(\phi)} &= 0
\end{aligned}
$$

Propiedades de la función $b(\theta)$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left\{\frac{\partial}{\partial\theta}\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]\right\} &= 0 \\
\mathbb{E}\left[\frac{y - b^{'}(\theta)}{a(\phi)}\right] &= 0 \\
\frac{\mathbb{E}(y) - b^{'}(\theta)}{a(\phi)} &= 0 \\
\frac{\mu - b^{'}(\theta)}{a(\phi)} &= 0
\end{aligned}$$

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left\{\frac{\partial}{\partial\theta}\left[\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right]\right\} &= 0 \\
\mathbb{E}\left[\frac{y-b^{'}(\theta)}{a(\phi)}\right] &= 0 \\
\frac{\mathbb{E}(y)-b^{'}(\theta)}{a(\phi)} &= 0 \\
\frac{\mu-b^{'}(\theta)}{a(\phi)} &= 0
\end{aligned}
$$

De donde se obtiene que $b^{'}(\theta)=\mu=\mathbb{E}(Y)$.

Ahora si, asumiendo que $f_Y(y|\theta,\phi)$ pertenece a la familia exponencial:

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left[\frac{\partial \log f_Y(y|\theta,\phi)}{\partial\theta}\right] &= 0 \\
\mathbb{E}\left\{\frac{\partial}{\partial\theta}\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]\right\} &= 0 \\
\mathbb{E}\left[\frac{y - b^{'}(\theta)}{a(\phi)}\right] &= 0 \\
\frac{\mathbb{E}(y) - b^{'}(\theta)}{a(\phi)} &= 0 \\
\frac{\mu - b^{'}(\theta)}{a(\phi)} &= 0
\end{aligned}$$

De donde se obtiene que $b^{'}(\theta) = \mu = \mathbb{E}(Y)$.

Con un razonamiento similar (que puede o no quedar como ejercicio) se puede demostrar que $\mathbb{V}ar(Y) = a(\phi)b^{''}(\theta)$.

Estimación

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud.

Estimación

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud. Antes de comenzar, recordemos la estructura de estos modelos.

Estimación

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud. Antes de comenzar, recordemos la estructura de estos modelos.

Modelo Lineal Generalizado

- $Y_1, Y_2, \ldots, Y_n$ son V.A. i.i.d. con densidad $f_Y(y|\theta, \phi)$ perteneciente a la flia. exponencial.
- $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$.
- $g(\mu) = \eta$.

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud. Antes de comenzar, recordemos la estructura de estos modelos.

Modelo Lineal Generalizado

- $Y_1, Y_2, \ldots, Y_n$ son V.A. i.i.d. con densidad $f_Y(y|\theta,\phi)$ perteneciente a la flia. exponencial.
- $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$.
- $g(\mu) = \eta$.

Adicionalmente, observamos que una manera de seleccionar la función de enlace era a través de la función de enlace canónica de la representación de $f_Y(y|\theta,\phi)$ como miembro de la familia exponencial, donde veíamos que $\theta = g(\mu)$.

Estimación

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud. Antes de comenzar, recordemos la estructura de estos modelos.

Modelo Lineal Generalizado

- $Y_1, Y_2, \ldots, Y_n$ son V.A. i.i.d. con densidad $f_Y(y|\theta, \phi)$ perteneciente a la flia. exponencial.
- $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$.
- $g(\mu) = \eta$.

Adicionalmente, observamos que una manera de seleccionar la función de enlace era a través de la función de enlace canónica de la representación de $f_Y(y|\theta, \phi)$ como miembro de la familia exponencial, donde veíamos que $\theta = g(\mu)$.

Entonces, **Si se opta por emplear la función de enlace canónica**

El método que emplearemos para estimar los parámetros de los *GLM* es máxima verosimilitud. Antes de comenzar, recordemos la estructura de estos modelos.

Modelo Lineal Generalizado

- $Y_1, Y_2, \ldots, Y_n$ son V.A. i.i.d. con densidad $f_Y(y|\theta, \phi)$ perteneciente a la flia. exponencial.
- $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$.
- $g(\mu) = \eta$.

Adicionalmente, observamos que una manera de seleccionar la función de enlace era a través de la función de enlace canónica de la representación de $f_Y(y|\theta, \phi)$ como miembro de la familia exponencial, donde veíamos que $\theta = g(\mu)$.

Entonces, **Si se opta por emplear la función de enlace canónica** $\theta = g(\mu) = \eta$.

Estimación

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

Estimación

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

O equivalentemente, debemos *anular* su gradiente (vector de derivadas de primer orden).

Esto es, encontrar el vector que anula el *score*.

$$\underbrace{\frac{\partial}{\partial\theta}\mathcal{L}(\theta)}_{s(\theta)} = \sum_i \underbrace{\frac{\partial}{\partial\theta}\ell(y_i|\theta, \phi)}_{s_i(\theta)} = 0$$

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

O equivalentemente, debemos *anular* su gradiente (vector de derivadas de primer orden).

Esto es, encontrar el vector que anula el *score*.

$$\underbrace{\frac{\partial}{\partial\theta}\mathcal{L}(\theta)}_{s(\theta)} = \sum_i \underbrace{\frac{\partial}{\partial\theta}\ell(y_i|\theta, \phi)}_{s_i(\theta)} = 0$$

Sin embargo, esto nos brindaría la estimación de θ siendo que nos interesa la de β.

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

O equivalentemente, debemos *anular* su gradiente (vector de derivadas de primer orden).

Esto es, encontrar el vector que anula el *score*.

$$\underbrace{\frac{\partial}{\partial \theta}\mathcal{L}(\theta)}_{s(\theta)} = \sum_i \underbrace{\frac{\partial}{\partial \theta}\ell(y_i|\theta, \phi)}_{s_i(\theta)} = 0$$

Sin embargo, esto nos brindaría la estimación de θ siendo que nos interesa la de β. Para resolver este problema, aplicamos la *regla de la cadena* y obtenemos.

$$\frac{\partial}{\partial \beta}\mathcal{L}(\theta) = \frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \beta} = 0$$

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

O equivalentemente, debemos *anular* su gradiente (vector de derivadas de primer orden).

Esto es, encontrar el vector que anula el *score*.

$$\underbrace{\frac{\partial}{\partial\theta}\mathcal{L}(\theta)}_{s(\theta)} = \sum_i \underbrace{\frac{\partial}{\partial\theta}\ell(y_i|\theta, \phi)}_{s_i(\theta)} = 0$$

Sin embargo, esto nos brindaría la estimación de θ siendo que nos interesa la de β. Para resolver este problema, aplicamos la *regla de la cadena* y obtenemos.

$$\begin{aligned} \frac{\partial}{\partial\beta}\mathcal{L}(\theta) &= \frac{\partial}{\partial\theta}\mathcal{L}(\theta)\frac{\partial\theta}{\partial\beta} = 0 \\ &= \frac{\partial}{\partial\theta}\mathcal{L}(\theta)\frac{\partial\theta}{\partial\mu}\frac{\partial\mu}{\partial\beta} = 0 \end{aligned}$$

El método de máxima verosimilitud plantea que debemos maximizar:

$$\mathcal{L}(\theta) = \sum_i \ell(y_i|\theta, \phi)$$

O equivalentemente, debemos *anular* su gradiente (vector de derivadas de primer orden).

Esto es, encontrar el vector que anula el *score*.

$$\underbrace{\frac{\partial}{\partial \theta}\mathcal{L}(\theta)}_{s(\theta)} = \sum_i \underbrace{\frac{\partial}{\partial \theta}\ell(y_i|\theta, \phi)}_{s_i(\theta)} = 0$$

Sin embargo, esto nos brindaría la estimación de θ siendo que nos interesa la de β. Para resolver este problema, aplicamos la *regla de la cadena* y obtenemos.

$$\begin{aligned}
\frac{\partial}{\partial \beta}\mathcal{L}(\theta) &= \frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \beta} = 0 \\
&= \frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \beta} = 0 \\
&= \frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta} = 0
\end{aligned}$$

Retomando la diapositiva anterior, nos interesa encontrar el β que resuelva.

$$\frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta} = 0$$

Estimación

Retomando la diapositiva anterior, nos interesa encontrar el β que resuelva.

$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta)\frac{\partial\theta}{\partial\mu}\frac{\partial\mu}{\partial\eta}\frac{\partial\eta}{\partial\beta} = 0$$

Es posible obtener expresiones para algunos de estos términos. Comencemos con el primero.

Retomando la diapositiva anterior, nos interesa encontrar el β que resuelva.

$$\frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta} = 0$$

Es posible obtener expresiones para algunos de estos términos. Comencemos con el primero.

$$\begin{aligned}
\tfrac{\partial}{\partial \theta}\mathcal{L}(\theta) &= \tfrac{\partial}{\partial \theta}\textstyle\sum_i \ell(y_i|\theta,\phi) \\
&= \tfrac{\partial}{\partial \theta}\textstyle\sum_i \left[\tfrac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i,\phi)\right] \\
&= \textstyle\sum_i \tfrac{y_i - b'(\theta)}{a(\phi)}
\end{aligned}$$

Retomando la diapositiva anterior, nos interesa encontrar el β que resuelva.

$$\frac{\partial}{\partial \theta}\mathcal{L}(\theta)\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta} = 0$$

Es posible obtener expresiones para algunos de estos términos. Comencemos con el primero.

$$\begin{aligned}
\tfrac{\partial}{\partial \theta}\mathcal{L}(\theta) &= \tfrac{\partial}{\partial \theta}\textstyle\sum_i \ell(y_i|\theta,\phi) \\
&= \tfrac{\partial}{\partial \theta}\textstyle\sum_i \left[\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i,\phi)\right] \\
&= \textstyle\sum_i \frac{y_i - b'(\theta)}{a(\phi)}
\end{aligned}$$

El último también es fácil:

$$\begin{aligned}
\tfrac{\partial \eta}{\partial \beta} &= \tfrac{\partial}{\partial \beta}x'\beta \\
\tfrac{\partial \eta}{\partial \beta} &= x'
\end{aligned}$$

Por último, **si se usa el enlace canónico**, se cumple que $\theta = \eta$, entonces:

Por último, **si se usa el enlace canónico**, se cumple que $\theta = \eta$, entonces:

$$
\begin{aligned}
\frac{\partial \theta}{\partial \mu} &= \frac{\partial \eta}{\partial \mu} \\
\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta} &= 1
\end{aligned}
$$

Estimación

Por último, **si se usa el enlace canónico**, se cumple que $\theta = \eta$, entonces:

$$\begin{aligned} \frac{\partial \theta}{\partial \mu} &= \frac{\partial \eta}{\partial \mu} \\ \frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta} &= 1 \end{aligned}$$

Finalmente, el problema de estimación se reduce a resolver:

$$s(\theta) = X^{'}\left[\frac{Y - b^{'}(\theta)}{a(\phi)}\right] = 0$$

El cual es un sistema de $k+1$ ecuaciones. Si queremos hacer *"aparecer"* a β, podemos recordar que $b^{'}(\theta) = \mu$, que $\mu = g^{-1}(\eta)$ y que $\eta = X\beta$

$$s(\beta) = X^{'}\left[\frac{Y - g^{-1}(X\beta)}{a(\phi)}\right] = 0$$

Estimación

Por último, **si se usa el enlace canónico**, se cumple que $\theta = \eta$, entonces:

$$\begin{aligned} \frac{\partial \theta}{\partial \mu} &= \frac{\partial \eta}{\partial \mu} \\ \frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta} &= 1 \end{aligned}$$

Finalmente, el problema de estimación se reduce a resolver:

$$s(\theta) = X^{'}\left[\frac{Y - b^{'}(\theta)}{a(\phi)}\right] = 0$$

El cual es un sistema de $k+1$ ecuaciones. Si queremos hacer *"aparecer"* a β, podemos recordar que $b^{'}(\theta) = \mu$, que $\mu = g^{-1}(\eta)$ y que $\eta = X\beta$

$$s(\beta) = X^{'}\left[\frac{Y - g^{-1}(X\beta)}{a(\phi)}\right] = 0$$

Sin embargo, estas ecuaciones solo admiten una solución cerrada en el caso en el que $Y_i \sim N(\mu_i, \sigma^2)$. Para los demás casos, es necesario trabajar un poco más.

Fisher scoring

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*.

Fisher scoring

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*. Este método consiste en anular el *score* de manera iterativa.

Fisher scoring

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*. Este método consiste en anular el *score* de manera iterativa. Para ello, es necesario plantear un desarrollo de primer orden del score.

$$s(\beta) \approx s(\beta_0) + \nabla s(\beta)|_{\beta_0}(\beta - \beta_0)$$

Fisher scoring

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*. Este método consiste en anular el *score* de manera iterativa. Para ello, es necesario plantear un desarrollo de primer orden del score.

$$s(\beta) \approx s(\beta_0) + \nabla s(\beta)|_{\beta_0}(\beta - \beta_0)$$

Luego, si evaluamos esta expresión en β^* (hipotética solución del sistema de ecuaciones), se cumple que $s(\beta^*) = 0$. Operando se obtiene:

Fisher scoring

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*. Este método consiste en anular el *score* de manera iterativa. Para ello, es necesario plantear un desarrollo de primer orden del score.

$$s(\beta) \approx s(\beta_0) + \nabla s(\beta)|_{\beta_0}(\beta - \beta_0)$$

Luego, si evaluamos esta expresión en β^* (hipotética solución del sistema de ecuaciones), se cumple que $s(\beta^*) = 0$. Operando se obtiene:

$$\beta^* \approx \beta_0 - \left[\nabla s(\beta)|_{\beta_0}\right]^{-1} s(\beta_0)$$

Siendo $\nabla s(\beta)|_{\beta_0}$ la matriz de información de Fisher $\mathcal{I}(\beta_0)$. Así, el esquema iterativo es:

El método de cabecera para estimar los parámetros de un *GLM* es *Fisher scoring*. Este método consiste en anular el *score* de manera iterativa. Para ello, es necesario plantear un desarrollo de primer orden del score.

$$s(\beta) \approx s(\beta_0) + \nabla s(\beta)|_{\beta_0}(\beta - \beta_0)$$

Luego, si evaluamos esta expresión en β^* (hipotética solución del sistema de ecuaciones), se cumple que $s(\beta^*) = 0$. Operando se obtiene:

$$\beta^* \approx \beta_0 - \left[\nabla s(\beta)|_{\beta_0}\right]^{-1} s(\beta_0)$$

Siendo $\nabla s(\beta)|_{\beta_0}$ la matriz de información de Fisher $\mathcal{I}(\beta_0)$. Así, el esquema iterativo es:

$$\beta^{(k+1)} = \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)})$$

Fisher scoring

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

Fisher scoring

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$\mathcal{I}(\beta) \quad = \quad \frac{\partial}{\partial \beta} \mathbf{s}(\beta)$$

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$\begin{aligned}\mathcal{I}(\beta) &= \tfrac{\partial}{\partial\beta} s(\beta) \\ &= \tfrac{\partial}{\partial\beta}\left\{X'\left[\tfrac{Y-b'(\theta)}{a(\phi)}\right]\right\} =\end{aligned}$$

Fisher scoring

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$
\begin{aligned}
\mathcal{I}(\beta) &= \tfrac{\partial}{\partial \beta} s(\beta) \\
&= \tfrac{\partial}{\partial \beta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\} = \tfrac{\partial}{\partial \theta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\}\tfrac{\partial \theta}{\partial \beta}
\end{aligned}
$$

Fisher scoring

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$\begin{aligned} \mathcal{I}(\beta) &= \frac{\partial}{\partial \beta} s(\beta) \\ &= \frac{\partial}{\partial \beta}\left\{X^{'}\left[\frac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\} = \frac{\partial}{\partial \theta}\left\{X^{'}\left[\frac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\}\frac{\partial \theta}{\partial \beta} \\ &= X^{'}\left[\frac{-b^{''}(\theta)}{a(\phi)}\right]\frac{\partial X\beta}{\partial \beta} = \end{aligned}$$

Fisher scoring

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$
\begin{aligned}
\mathcal{I}(\beta) &= \tfrac{\partial}{\partial \beta} s(\beta) \\
&= \tfrac{\partial}{\partial \beta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\} = \tfrac{\partial}{\partial \theta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\}\tfrac{\partial \theta}{\partial \beta} \\
&= X^{'}\left[\tfrac{-b^{''}(\theta)}{a(\phi)}\right]\tfrac{\partial X\beta}{\partial \beta} = X^{'}\left[\tfrac{-b^{''}(\theta)}{a(\phi)}\right]X
\end{aligned}
$$

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$\begin{aligned}\mathcal{I}(\beta) &= \tfrac{\partial}{\partial\beta} s(\beta) \\ &= \tfrac{\partial}{\partial\beta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\} = \tfrac{\partial}{\partial\theta}\left\{X^{'}\left[\tfrac{Y-b^{'}(\theta)}{a(\phi)}\right]\right\}\tfrac{\partial\theta}{\partial\beta} \\ &= X^{'}\left[\tfrac{-b^{''}(\theta)}{a(\phi)}\right]\tfrac{\partial X\beta}{\partial\beta} = X^{'}\left[\tfrac{-b^{''}(\theta)}{a(\phi)}\right]X\end{aligned}$$

Para *"deshacernos"* de θ debemos recordar que $a(\phi)b^{''}(\theta) = \mathbb{V}ar(Y)$ y que por ende $\frac{b^{''}(\theta)}{a(\phi)} = \frac{\mathbb{V}ar(Y)}{a^2(\phi)}$. Entonces:

Para terminar, necesitamos una expresión para $\mathcal{I}(\beta)$ en la expresión anterior:

$$\begin{aligned}
\mathcal{I}(\beta) &= \tfrac{\partial}{\partial \beta} s(\beta) \\
&= \tfrac{\partial}{\partial \beta} \left\{ X^{'} \left[\tfrac{Y - b^{'}(\theta)}{a(\phi)} \right] \right\} = \tfrac{\partial}{\partial \theta} \left\{ X^{'} \left[\tfrac{Y - b^{'}(\theta)}{a(\phi)} \right] \right\} \tfrac{\partial \theta}{\partial \beta} \\
&= X^{'} \left[\tfrac{-b^{''}(\theta)}{a(\phi)} \right] \tfrac{\partial X \beta}{\partial \beta} = X^{'} \left[\tfrac{-b^{''}(\theta)}{a(\phi)} \right] X
\end{aligned}$$

Para *"deshacernos"* de θ debemos recordar que $a(\phi) b^{''}(\theta) = \mathbb{V}ar(Y)$ y que por ende $\frac{b^{''}(\theta)}{a(\phi)} = \frac{\mathbb{V}ar(Y)}{a^2(\phi)}$. Entonces:

$$\mathcal{I}(\beta) = -X^{'} \left[\frac{\mathbb{V}ar(Y)}{a^2(\phi)} \right] X$$

Téngase en cuenta que la expresión de $\mathbb{V}ar(Y)$ NO es la varianza muestral sino que depende del modelo GLM y por ende también es función de β. Por lo tanto:

$$\mathcal{I}(\beta) = -X^{'} W(\beta) X$$

Fisher scoring

Reuniendo todos los elementos en la ecuación del método iterativo:

Fisher scoring

Reuniendo todos los elementos en la ecuación del método iterativo:

$$
\begin{aligned}
\beta^{(k+1)} &= \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)}) \\
\beta^{(k+1)} &= \beta^{(k)} + \left[X' W(\beta^{(k)}) X\right]^{-1} X' \left[\frac{Y - g^{-1}(X\beta^{(k)})}{a(\phi)}\right]
\end{aligned}
$$

Reuniendo todos los elementos en la ecuación del método iterativo:

$$
\begin{aligned}
\beta^{(k+1)} &= \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)}) \\
\beta^{(k+1)} &= \beta^{(k)} + \left[X' W(\beta^{(k)}) X\right]^{-1} X' \left[\frac{Y - g^{-1}(X\beta^{(k)})}{a(\phi)}\right]
\end{aligned}
$$

Si reescribimos $\beta^{(k)} = \left[X' W(\beta) X\right]^{-1} \left[X' W(\beta) X\right] \beta^{(k)}$:

Reuniendo todos los elementos en la ecuación del método iterativo:

$$\begin{aligned}\beta^{(k+1)} &= \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)}) \\ \beta^{(k+1)} &= \beta^{(k)} + \left[X' W(\beta^{(k)}) X\right]^{-1} X' \left[\frac{Y - g^{-1}(X\beta^{(k)})}{a(\phi)}\right]\end{aligned}$$

Si reescribimos $\beta^{(k)} = \left[X' W(\beta) X\right]^{-1} \left[X' W(\beta) X\right] \beta^{(k)}$:

$$\beta^{(k+1)} = \frac{1}{a(\phi)} \left[X' W(\beta^{(k)}) X\right]^{-1} X' W(\beta^{(k)}) \underbrace{\left\{X' \beta^{(k)} - W^{-1}(\beta^{(k)}) \left[Y - g^{-1}(X\beta^{(k)})\right]\right\}}$$

Reuniendo todos los elementos en la ecuación del método iterativo:

$$\begin{aligned}\beta^{(k+1)} &= \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)}) \\ \beta^{(k+1)} &= \beta^{(k)} + \left[X' W(\beta^{(k)}) X\right]^{-1} X' \left[\frac{Y - g^{-1}(X\beta^{(k)})}{a(\phi)}\right]\end{aligned}$$

Si reescribimos $\beta^{(k)} = \left[X' W(\beta) X\right]^{-1} \left[X' W(\beta) X\right] \beta^{(k)}$:

$$\beta^{(k+1)} = \frac{1}{a(\phi)} \left[X' W(\beta^{(k)}) X\right]^{-1} X' W(\beta^{(k)}) \underbrace{\left\{X' \beta^{(k)} - W^{-1}(\beta^{(k)}) \left[Y - g^{-1}(X\beta^{(k)})\right]\right\}}_{z(\beta^{(k)})}$$

Reuniendo todos los elementos en la ecuación del método iterativo:

$$\begin{aligned}\beta^{(k+1)} &= \beta^{(k)} - \mathcal{I}(\beta^{(k)})^{-1} s(\beta^{(k)}) \\ \beta^{(k+1)} &= \beta^{(k)} + \left[X' W(\beta^{(k)}) X\right]^{-1} X' \left[\frac{Y - g^{-1}(X\beta^{(k)})}{a(\phi)}\right]\end{aligned}$$

Si reescribimos $\beta^{(k)} = \left[X' W(\beta) X\right]^{-1} \left[X' W(\beta) X\right] \beta^{(k)}$:

$$\beta^{(k+1)} = \frac{1}{a(\phi)} \left[X' W(\beta^{(k)}) X\right]^{-1} X' W(\beta^{(k)}) \underbrace{\left\{X' \beta^{(k)} - W^{-1}(\beta^{(k)}) \left[Y - g^{-1}(X\beta^{(k)})\right]\right\}}_{z(\beta^{(k)})}$$

Donde el término $z(\beta^{(k)})$ es comunmente llamado *working response*.

Fisher scoring

Observemos que el estimador (iterativo) de β:

Fisher scoring

Observemos que el estimador (iterativo) de β:

$$\beta^{(k+1)} \quad = \quad \frac{1}{a(\phi)}\left[X^{'}W(\beta^{(k)})X\right]^{-1}X^{'}W(\beta^{(k)})z(\beta^{(k)})$$

es la solución del problema de **mínimos cuadrados ponderados iterativos** (IRWLS).

Fisher scoring

Observemos que el estimador (iterativo) de β:

$$\beta^{(k+1)} = \frac{1}{a(\phi)}\left[X'W(\beta^{(k)})X\right]^{-1}X'W(\beta^{(k)})z(\beta^{(k)})$$

es la solución del problema de **mínimos cuadrados ponderados iterativos** (IRWLS).

La forma en la que procede este algoritmo en cada iteración es la siguiente:

1. Disponiendo del iterando $\beta^{(k)}$ se calcula:
   - $W(\beta^{(k)})$
   - $\eta^{(k)} = X'\beta^{(k)}$
   - $\mu(\beta^{(k)}) = g^{-1}(\eta^{(k)})$ y finalmente se calcula la *working response*
   - $z(\beta^{(k)}) = \eta^{(k)} - W^{-1}(\beta^{(k)})\left[Y - \mu\left(\beta^{(k)}\right)\right]$
2. Disponiendo de los valores anteriores, se actualiza el vector de estimaciones, con

$$\beta^{(k+1)} = \frac{1}{a(\phi)}\left[X'W(\beta^{(k)})X\right]^{-1}X'W(\beta^{(k)})z(\beta^{(k)})$$

3. Se vuelve al paso 1 hasta alcanzar convergencia (por ejemplo $||s(\beta^{(k)})|| < tol$)

Covarianza de $\hat{\beta}$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Covarianza de $\hat{\beta}$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Luego de alcanzar convergencia, es posible obtener una expresión para esta matriz de covarianzas (en la siguiente expresión se quitan los superíndices (k) y la dependencia de W del vector β).

$$\hat{\beta} = \frac{1}{a(\phi)} \left[X' WX\right]^{-1} X' W \left[\eta - W^{-1}(Y - \mu)\right]$$

El único término aleatorio en esta expresión es Y dentro de la *working response*. Por ende:

Covarianza de $\hat{\beta}$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Luego de alcanzar convergencia, es posible obtener una expresión para esta matriz de covarianzas (en la siguiente expresión se quitan los superíndices (k) y la dependencia de W del vector β).

$$\hat{\beta} = \frac{1}{a(\phi)} \left[X^{'} WX\right]^{-1} X^{'} W \left[\eta - W^{-1}(Y - \mu)\right]$$

El único término aleatorio en esta expresión es Y dentro de la *working response*. Por ende:

$$\mathbb{V}ar(\hat{\beta}) \quad = \quad \mathbb{V}ar \left\{ \tfrac{1}{a(\phi)} \left[X^{'} WX\right]^{-1} X^{'} WW^{-1} Y \right\}$$

Covarianza de $\hat{\beta}$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Luego de alcanzar convergencia, es posible obtener una expresión para esta matriz de covarianzas (en la siguiente expresión se quitan los superíndices (k) y la dependencia de W del vector β).

$$\hat{\beta} = \frac{1}{a(\phi)} \left[X^{'} WX\right]^{-1} X^{'} W \left[\eta - W^{-1}(Y - \mu)\right]$$

El único término aleatorio en esta expresión es Y dentro de la *working response*. Por ende:

$$\begin{aligned} \mathbb{V}ar(\hat{\beta}) &= \mathbb{V}ar\left\{\tfrac{1}{a(\phi)} \left[X^{'} WX\right]^{-1} X^{'} WW^{-1} Y\right\} \\ &= \tfrac{1}{a(\phi)} \left[X^{'} WX\right]^{-1} X^{'} \mathbb{V}ar(Y) \tfrac{1}{a(\phi)} \left\{\left[X^{'} WX\right]^{-1} X^{'}\right\}^{'} \end{aligned}$$

Covarianza de $\hat{\beta}$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Luego de alcanzar convergencia, es posible obtener una expresión para esta matriz de covarianzas (en la siguiente expresión se quitan los superíndices (k) y la dependencia de W del vector β).

$$\hat{\beta} = \frac{1}{a(\phi)}\left[X^{'}WX\right]^{-1}X^{'}W\left[\eta - W^{-1}(Y-\mu)\right]$$

El único término aleatorio en esta expresión es Y dentro de la *working response*. Por ende:

$$\begin{aligned}
\mathbb{V}ar(\hat{\beta}) &= \mathbb{V}ar\left\{\tfrac{1}{a(\phi)}\left[X^{'}WX\right]^{-1}X^{'}WW^{-1}Y\right\} \\
&= \tfrac{1}{a(\phi)}\left[X^{'}WX\right]^{-1}X^{'}\mathbb{V}ar(Y)\tfrac{1}{a(\phi)}\left\{\left[X^{'}WX\right]^{-1}X^{'}\right\}^{'} \\
&= \left[X^{'}WX\right]^{-1}X^{'}\underbrace{W}_{\frac{\mathbb{V}ar(Y)}{a^2(\phi)}}X\left[X^{'}WX\right]^{-1}
\end{aligned}$$

Una vez estimado el vector de parámetros suele resultar de interés conocer su precisión.

Luego de alcanzar convergencia, es posible obtener una expresión para esta matriz de covarianzas (en la siguiente expresión se quitan los superíndices (k) y la dependencia de W del vector β).

$$\hat{\beta} = \frac{1}{a(\phi)} \left[X' WX\right]^{-1} X' W \left[\eta - W^{-1}(Y - \mu)\right]$$

El único término aleatorio en esta expresión es Y dentro de la *working response*. Por ende:

$$\begin{aligned}
\mathbb{V}ar(\hat{\beta}) &= \mathbb{V}ar\left\{ \tfrac{1}{a(\phi)} \left[X' WX\right]^{-1} X' WW^{-1} Y \right\} \\
&= \tfrac{1}{a(\phi)} \left[X' WX\right]^{-1} X' \mathbb{V}ar(Y) \tfrac{1}{a(\phi)} \left\{ \left[X' WX\right]^{-1} X' \right\}' \\
&= \left[X' WX\right]^{-1} X' \underbrace{W}_{\frac{\mathbb{V}ar(Y)}{a^2(\phi)}} X \left[X' WX\right]^{-1} \\
&= \left[X' WX\right]^{-1} = -\mathcal{I}(\beta)^{-1}
\end{aligned}$$

Predicción

Luego de estimados los parámetros del modelo, el interés puede recaer sobre la realización de predicciones.

Predicción

Luego de estimados los parámetros del modelo, el interés puede recaer sobre la realización de predicciones.

Si recordamos la definición de la función de enlace, la predicción de una futura observación se realizaría como:

$$\hat{y} = \mathbb{E}(Y|X) = g^{-1}(X\hat{\beta})$$

Predicción

Luego de estimados los parámetros del modelo, el interés puede recaer sobre la realización de predicciones.

Si recordamos la definición de la función de enlace, la predicción de una futura observación se realizaría como:

$$\hat{y} = \mathbb{E}(Y|X) = g^{-1}(X\hat{\beta})$$

En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{E}(Y|X) = X\hat{\beta}$

Predicción

Luego de estimados los parámetros del modelo, el interés puede recaer sobre la realización de predicciones.

Si recordamos la definición de la función de enlace, la predicción de una futura observación se realizaría como:

$$\hat{y} = \mathbb{E}(Y|X) = g^{-1}(X\hat{\beta})$$

En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{E}(Y|X) = X\hat{\beta}$
- **Bernoulli**: $\mathbb{E}(Y|X) = \frac{e^{X\hat{\beta}}}{1+e^{X\hat{\beta}}}$

Predicción

Luego de estimados los parámetros del modelo, el interés puede recaer sobre la realización de predicciones.

Si recordamos la definición de la función de enlace, la predicción de una futura observación se realizaría como:

$$\hat{y} = \mathbb{E}(Y|X) = g^{-1}(X\hat{\beta})$$

En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{E}(Y|X) = X\hat{\beta}$
- **Bernoulli**: $\mathbb{E}(Y|X) = \frac{e^{X\hat{\beta}}}{1+e^{X\hat{\beta}}}$
- **Poisson**: $\mathbb{E}(Y|X) = e^{X\hat{\beta}}$

Desvío de la predicción

En el modelo lineal, vimos que HY y que $\mathbb{V}ar(\hat{Y}) = \sigma^2 H$.

Desvío de la predicción

En el modelo lineal, vimos que HY y que $\mathbb{V}ar(\hat{Y}) = \sigma^2 H$. En el ámbito de los *GLM* esto no es tan inmediato debido a que $\mathbb{E}(Y) = g^{-1}(X\beta)$ siendo $g(.)$ típicamente una función no lineal.

En el modelo lineal, vimos que HY y que $\mathbb{V}ar(\hat{Y}) = \sigma^2 H$. En el ámbito de los *GLM* esto no es tan inmediato debido a que $\mathbb{E}(Y) = g^{-1}(X\beta)$ siendo $g(.)$ típicamente una función no lineal.

Una herramienta que nos puede ayudar a solucionar este problema es el *método delta*

MÉTODO DELTA

Sea $\hat{\theta}$ un estimador con esperanza θ_0 con la siguiente distribución asintótica:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \rightarrow N(0, V)$$

Desvío de la predicción

En el modelo lineal, vimos que HY y que $\mathbb{V}ar(\hat{Y}) = \sigma^2 H$. En el ámbito de los *GLM* esto no es tan inmediato debido a que $\mathbb{E}(Y) = g^{-1}(X\beta)$ siendo $g(.)$ típicamente una función no lineal.

Una herramienta que nos puede ayudar a solucionar este problema es el *método delta*

Método delta

Sea $\hat{\theta}$ un estimador con esperanza θ_0 con la siguiente distribución asintótica:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to N(0, V)$$

Luego, siendo $g(.)$ una función continua, se pude demostrar que:

$$\sqrt{n}\left(g(\hat{\theta}) - g(\theta_0)\right) \to N(0, g^{'}(\theta_0)^2 V)$$

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$.

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$.

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.
2. Evaluarla en $\hat{\eta}$ (porque es lo más parecido que tenemos a η).

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.
2. Evaluarla en $\hat{\eta}$ (porque es lo más parecido que tenemos a η).
3. Multiplicar eso por $\hat{\sigma}^2_{\hat{\eta}}$

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.
2. Evaluarla en $\hat{\eta}$ (porque es lo más parecido que tenemos a η).
3. Multiplicar eso por $\hat{\sigma}^2_{\hat{\eta}}$

En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{V}ar(\hat{Y}) = \hat{\sigma}^2_{\hat{\eta}}$

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X^{'}$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.
2. Evaluarla en $\hat{\eta}$ (porque es lo más parecido que tenemos a η).
3. Multiplicar eso por $\hat{\sigma}^2_{\hat{\eta}}$

En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{V}ar(\hat{Y}) = \hat{\sigma}^2_{\hat{\eta}}$
- **Bernoulli**: $\mathbb{V}ar(\hat{Y}) = \left[\frac{e^{\hat{\eta}}}{\left(1+e^{\hat{\eta}}\right)^2}\right]^2 \hat{\sigma}^2_{\hat{\eta}}$

Desvío de la predicción

En nuestro caso $\hat{\eta} = X\hat{\beta}$, cuya varianza es $-X\mathcal{I}(\beta)^{-1}X'$. En términos prácticos, al sustituir β por $\hat{\beta}$, tenemos una estimación de esa varianza a la cual llamaremos $\hat{\sigma}^2_{\hat{\eta}}$.

Y ahora si aplicamos el método delta debido a que para obtener las predicciones, usamos el hecho de que $\hat{Y} = g^{-1}(\hat{\eta})$. Para obtener la varianza de las predicciones debemos:

1. Derivar $g^{-1}(.)$ y elevarla al cuadrado.
2. Evaluarla en $\hat{\eta}$ (porque es lo más parecido que tenemos a η).
3. Multiplicar eso por $\hat{\sigma}^2_{\hat{\eta}}$

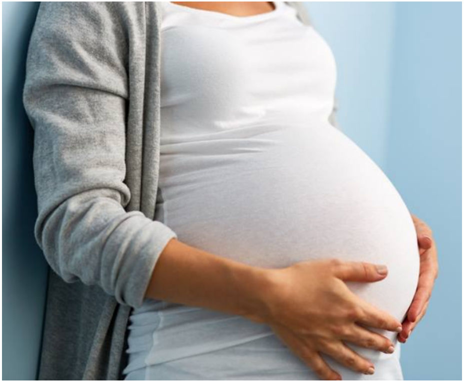
En los *GLM* que vimos la clase pasada esto sería:

- **Normal**: $\mathbb{V}ar(\hat{Y}) = \hat{\sigma}^2_{\hat{\eta}}$
- **Bernoulli**: $\mathbb{V}ar(\hat{Y}) = \left[\frac{e^{\hat{\eta}}}{\left(1+e^{\hat{\eta}}\right)^2}\right]^2 \hat{\sigma}^2_{\hat{\eta}}$
- **Poisson**: $\mathbb{V}ar(\hat{Y}) = \left[e^{\hat{\eta}}\right]^2 \hat{\sigma}^2_{\hat{\eta}}$

Ejemplo

Volvamos a los datos del bajo peso al nacer y prestemos atención al proceso de estimación y predicción.

Volvamos a los datos del bajo peso al nacer y prestemos atención al proceso de estimación y predicción.



Volvamos al R.

En la próxima clase

La próxima hablaremos de:

- Aspectos inferenciales de los *GLM*.
- Específicamente del modelo de regresión logística.
- Veremos el problema de la *separabilidad*.
- Mencionaremos la posibilidad de utilizar funciones de enlace diferentes a la canónica.

Carmona, Francesc (2003). *Modelos Lineales (notas de curso)*. Departament d'Estadística.

Farraway, Julian (2014). *Linear Models with R, second edition*. Chapman Hall/CRC.

McCullagh, P. y J.A. Nelder (1983). *Generalized Linear Models*. Chapman Hall/CRC.

Rencher, Alvin y Bruce Schaalje (2008). *Linear Models in Statistics, second edition*. John Wiley Sons, Inc.

¿Preguntas?

Muchas Gracias