INFERÊNCIA EM MODELOS LINEARES GENERALIZADOS

ANÁLISE DE DEVIANCE

- A análise de deviance é uma generalização, para modelos lineares generalizados, da análise de variância.

- No caso de modelos lineares, utiliza-se a chamada "extra soma de quadrados" para avaliar a significância de termos incluídos ao modelo;

- Em MLG, de forma semelhante, é de interesse testar a significância da inclusão de novos termos. Neste sentido, usaremos com frequência a expressão *modelos encaixados*;

- Dizemos que dois modelos são encaixados se um modelo é obtido a partir do outro impondo alguma restrição aos valores dos parâmetros (é usual assumir valor zero aos parâmetros, caso se deseje investigar a hipótese de nulidade dos mesmos);

- Na sequência são apresentados os preditores lineares de diferentes modelos lineares generalizados para avaliarmos se configuram modelos encaixados.

- Caso 1:

$$\begin{array}{ll} Modelo\,1 - & g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ Modelo\,2 - & g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_3 x_3 \end{array}$$

\- Modelos encaixados!

A comparação dois modelos apresentados no par 1 poderia fundamentar o teste da hipótese:

$$H_0 : \beta_2 = \beta_4 = 0,$$

contra a alternativa que os parâmetros sob teste não são conjuntamente nulos.

- Caso 2:

$$\begin{array}{ll} Modelo\,1- & g(\mu_{\mathbf{x}})=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4 \\ Modelo\,2- & g(\mu_{\mathbf{x}})=\beta_0+2\,x_1-x_3+\beta_4 x_4 \end{array}$$ - Modelos encaixados!

A comparação dois modelos apresentados no par 2 poderia fundamentar o teste da seguinte hipótese:

$$H_0:\beta_1=2;\beta_2=0;\beta_3=-1.$$

- Caso 3:

$$\begin{array}{ll} Modelo\,1 - & g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \\ Modelo\,2 - & g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \end{array}$$ - Modelos encaixados!

A comparação dois modelos apresentados no par 3 poderia fundamentar o teste da seguinte hipótese:

$$H_0 : \beta_3 = 0.$$

Repare que, mediante este par de hipóteses, estaríamos testando a existência de interação entre x_1 e x_2.

- Caso 4:

$$\begin{array}{ll} Modelo\,1- & g(\mu_{\mathbf{x}})=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\beta_3 x_1^3 \\ Modelo\,2- & g(\mu_{\mathbf{x}})=\beta_0+\beta_1 x_1 \end{array}$$ - Modelos encaixados!

A comparação dois modelos apresentados no par 4 poderia fundamentar o teste da seguinte hipótese:

$$H_0:\beta_2=0,\beta_3=0.$$

Repare que, mediante este par de hipóteses, estaríamos testando a existência de efeito cúbico ou quadrático de x_1 em y.

Nota – Note que em qualquer um dos quatro exemplos apresentados, a hipótese nula representa o modelo restrito e a hipótese alternativa o modelo não restrito. No contexto de teste de hipóteses, a rejeição de H_0 corresponde à diferença dos ajustes dos dois modelos, sendo que se deve optar, nesses casos, pelo modelo não restrito (com mais parâmetros).

- Caso 5:

$$\begin{aligned} &Modelo\,1 - \quad g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &Modelo\,2 - \quad g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_4 x_4 \end{aligned}$$

- Modelos não encaixados!

- Caso 6:

$$\begin{aligned} &Modelo\,1 - \quad g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &Modelo\,2 - \quad g(\mu_{\mathbf{x}}) = \beta_0 + \beta_1 x_1 + \beta_2 \ln(x_2) \end{aligned}$$

- Modelos não encaixados!

Teste da razão de verossimilhanças (TRV) em MLG

- O teste da razão de verossimilhanças é amplamente utilizado em MLG para testar a nulidade conjunta de (ou alguma outra restrição envolvendo) parâmetros de modelos lineares generalizados.

- Seja M_p um MLG com p parâmetros e M_q um modelo encaixado a M_p, a partir de uma restrição a $p-q$ parâmetros, restando $q<p$ parâmetros não fixados (irrestritos).

- Considere D_p e D_q, respectivamente, os desvios de M_p e M_q. A estatística é uma medida de diferença dos ajuste de M_p e M_q, que pode ser entendida como o ganho de ajuste decorrente da inclusão de $p-q$ parâmetros ao modelo mais simples.

- A estatística do teste da razão de verossimilhanças para comparação dos dois modelos fica dada por:

$$\xi_{RV} = \frac{D_q - D_p}{\phi} = 2\phi^{-1}\left\{l\left(\hat{\boldsymbol{\mu}}_p;\mathbf{y}\right) - l\left(\hat{\boldsymbol{\mu}}_q;\mathbf{y}\right)\right\} = 2\phi^{-1}\left\{\ln\left[\frac{L\left(\hat{\boldsymbol{\mu}}_p;\mathbf{y}\right)}{L\left(\hat{\boldsymbol{\mu}}_q;\mathbf{y}\right)}\right]\right\},$$

que, sob a hipótese nula de que as restrições são válidas, tem **assintoticamente** distribuição χ^2_{p-q}.

- Caso a hipótese nula seja de nulidade de $p-q$ parâmetros e o resultado do teste indique a não rejeição de H_0, isso pode justificar a eliminação dos termos (covariáveis, fatores...) associados aos $p-q$ parâmetros 'nulos'.

Nota – O teste da razão de verossimilhanças pode ser aplicado a um único parâmetro (Exemplo: $H_0 : \beta_k = 0$ *vs* $H_1 : \beta_k \neq 0$), sendo que neste caso, sob H_0 ξ_{RV} tem, assintoticamente, distribuição χ_1^2.

Nota – Podemos testar a significância do modelo ajustado considerando a hipótese nula $H_0 : \beta_1 = \beta_2 = ... = \beta_{p-1} = 0$, ou seja, comparando o ajuste do modelo com p parâmetros ao do modelo nulo (só com intercepto).

No R:

```
ajuste1=glm(...)### “Modelo maior”
ajuste2=glm(...)### “Modelo menor”
anova(ajuste2,ajuste1,test=’Chisq’)
```

Procedimento geral para o teste da razão de verossimilhanças em Modelos Lineares Generalizados

1. Formular as hipóteses de interesse e estabelecer adequadamente os modelos restrito (M_q) e não restrito (M_p) correspondentes;

2. Ajustar os dois modelos aos dados e extrair os correspondentes desvios (D_q e D_p);

3. Calcular a estatística do teste da razão de verossimilhanças (ξ_{RV});

4. Com base no valor de ξ_{RV}, testar a hipótese nula de que a restrição considerada é válida. Por exemplo, para um nível de significância α, rejeitamos H_0 se ξ_{RV} exceder o quantil $(1-\alpha)$ da distribuição χ^2_{p-q}.

Teste F para o caso em que ϕ é desconhecido

- Para as distribuições em que o parâmetro de dispersão é desconhecido (Normal, Gama e Normal Inversa, por exemplo), pode-se utilizar uma estimativa e considerar como alternativa o uso do teste F, ao invés de χ^2.

- A estatística do teste F é definida por:

$$\xi_{RV} = \frac{\left(D_q - D_p\right) / (p-q)}{D_p / (n-p)},$$

que, sob a hipótese nula de que as restrições impostas em H_0 são válidas, tem **assintoticamente** distribuição $F_{p-q,n-p}$.

Nota – Pode-se substituir $D_p/(n-p)$ no denominador da estatística F por alguma estimativa consistente de ϕ.

No R:

```
ajuste1=glm(...)### "Modelo maior"
ajuste2=glm(...)### "Modelo menor"
anova(ajuste2,ajuste1,test='F')
```

Análise de deviance desvio (Tabela ANODEV)

- A análise de deviance configura uma extensão da análise de variância para os modelos lineares generalizados.

- Baseia-se na comparação das deviances avaliadas para modelos encaixados, permitindo testar o efeito de sucessivas inclusões inclusão (ou exclusões) de variáveis, fatores e interações a um modelo corrente.

- A Tabela Anodev é a representação de uma sequência de TRVs para um modelo linear generalizado, em que os termos do preditor linear são acrescentados sucessivamente ao modelo (começando pelo modelo nulo), e a significância de suas inclusões avaliadas via TRV.

- A título de ilustração, considere um MLG qualquer, com quatro variáveis no preditor linear (X_1, X_2, X_3, X_4). Então, na tabela Anodev serão apresentados os desvios, as diferenças de deviances, os correspondentes graus de liberdade e os testes de razão de verossimilhança para:

    - Inclusão de X_1 ao modelo que contém apenas o intercepto;
    - Inclusão de X_2 ao modelo que contém X_1;
    - Inclusão de X_3 ao modelo que contém X_1 e X_2;
    - Inclusão de X_4 ao modelo que contém X_1, X_2 e X_3.

Notas-

1. A ordem de inclusão das variáveis é determinada pelo usuário e, exceto em casos bem específicos, vai alterar a significância das variáveis;

2. A ordem de inclusão de termos ao modelo, quando na ocorrência de interações, deve obedecer ao *principio hierárquico*. Ou seja, se temos no modelo $X_1, X_2 \; e \; X_1 \times X_2$, primeiramente inserimos ao modelo X_1 e X_2 (na ordem que bem se entender) para **depois** inserir o termo correspondente à interação. O mesmo vale para modelos polinomiais, em que os termos de menor ordem são os primeiros a serem inseridos.

No R: Comando `anova`.

- Uma forma alternativa de se fazer a análise do desvio é avaliando a significância de uma variável quando inserida ao modelo que contém todas as demais variáveis, exceto a variável em questão.

- A título de ilustração, considere um MLG qualquer, com quatro variáveis no preditor linear (X_1, X_2, X_3, X_4). Então, na tabela Anodev serão apresentados os desvios, as diferenças de deviances, os correspondentes graus de liberdade e os testes de razão de verossimilhança para:

    - Inclusão de X_1 ao modelo que contém X_2, X_3 e X_4;
    - Inclusão de X_2 ao modelo que contém X_1, X_3 e X_4;
    - Inclusão de X_3 ao modelo que contém X_1, X_2 e X_4;
    - Inclusão de X_4 ao modelo que contém X_1, X_2 e X_3.

No R: Comando `Anova`, pacote car.

Teste de Wald

- O teste de Wald baseia-se na distribuição assintótica normal dos estimadores de máxima verossimilhança dos parâmetros do modelo.

- Seja $\hat{\beta}_j$ o estimador de máxima verossimilhança de β_j, um particular parâmetro de um MLG. Conforme discutido anteriormente, para $n \to \infty$,

$$\hat{\beta}_j \sim Normal\left(\beta_j, Var\left(\hat{\beta}_j\right)\right),$$

em que $Var\left(\hat{\beta}_j\right)$ é estimada através do correspondente termo da diagonal da matriz de covariâncias $\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\hat{\phi}$. Vamos denotar por $ep\left(\hat{\beta}_j\right) = \sqrt{\widehat{Var}\left(\hat{\beta}_j\right)}$ o erro padrão de $\hat{\beta}_j$.

- Embora possam ser aplicados ao teste de hipóteses de dois ou mais parâmetros, o uso mais frequente do teste de Wald contempla apenas um parâmetro por vez. Em situações envolvendo mais parâmetros, é mais usual aplicar o teste da razão de verossimilhanças.

- Considere então o seguinte par de hipóteses:

$$\begin{array}{l} H_0 : \beta_j = \beta_j^{(0)} \\ H_1 : \beta_j \neq \beta_j^{(0)} \end{array},$$

em que $\beta_j^{(0)}$ é algum valor postulado para β_j (é comum tomarmos $\beta_j^{(0)} = 0$, a fim de testarmos a nulidade de β_j). Então, o teste de Wald baseia-se na seguinte estatística-teste:

$$Z_t = \frac{\hat{\beta}_j - \beta_j^{(0)}}{ep(\hat{\beta}_j)},$$

que, sob a hipótese nula, tem assintoticamente distribuição Normal padrão.

- Para um nível de significância α, rejeitaremos H_0 caso $|Z_t| > z_{1-\alpha/2}$, em $z_{1-\alpha/2}$ representa o quantil $1-\alpha/2$ da distribuição Normal padrão.

- Nos casos em que ϕ é desconhecido, pode-se usar a distribuição $t-Student$ com $n-p$ graus de liberdade, rejeitando H_0, para um nível de significância α, se $|Z_t| > t_{n-p;1-\alpha/2}$.

No R: A estatística e o teste de Wald são apresentados no próprio `summary` de um MLG.

Nota – A função `waldtest`, do pacote `lmtest` permite aplicar o teste de Wald para hipóteses envolvendo $p \geq 2$ parâmetros, baseado numa distribuição assintótica χ^2_{n-p}.

Intervalos de confiança

- Dentre os métodos disponíveis para obtenção de intervalos de confiança em Modelos Lineares Generalizados, serão destacados os intervalos baseados na razão de verossimilhanças e na estatística de Wald. Mais adiante discutiremos o uso de simulação (bootstrap) para a obtenção dos intervalos.

Intervalos de confiança baseados na razão de verossimilhanças

- Um intervalo com nível de confiança assintótico $1-\alpha$ para β_j , baseado na razão de verossimilhanças, contém todos os valores $\beta_j^{(0)}$ para os quais a hipótese nula $H_0 : \beta_j = \beta_j^{(0)}$ não seria rejeitada pelo TRV, ao nível de significância α.

- Para fins de ilustração, considerando um nível de confiança (assintótico) de 95%, o intervalo de confiança para β_j conteria todo $\beta_j^{(0)}$ para o qual a hipótese $H_0 : \beta_j = \beta_j^{(0)}$ produzisse:

$$\xi_{RV} = \frac{D_0 - D_1}{\phi} = 2\phi^{-1}\left\{\ln\left[\frac{L(\hat{\boldsymbol{\mu}}_1; \mathbf{y})}{L(\hat{\boldsymbol{\mu}}_0; \mathbf{y})}\right]\right\} \leq \chi^2_{0,95;1} = 3{,}84,$$

sendo D_0 o desvio avaliado considerando $\beta_j = \beta_j^{(0)}$ e D_1 o desvio avaliado no modelo sem restrição para β_j.

No R: Função `confint`.

Intervalos de confiança baseados na estatística de Wald

- Uma vez que, assintoticamente:

$$\frac{\hat{\beta}_j - \beta_j}{ep(\hat{\beta}_j)} \sim Normal(0,1),$$

pode-se determinar quantis $z_{\alpha/2}$ e $z_{1-\alpha/2}$ tais que:

$$P\left(z_{\alpha/2} < \frac{\hat{\beta}_j - \beta_j}{ep(\hat{\beta}_j)} < z_{1-\alpha/2} \right) \approx 1-\alpha,\ 0 < \alpha < 1.$$

- Isolando β_j no centro da desigualdade, temos:

$$P\left(\hat{\beta}_j - z_{1-\alpha/2} ep(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + z_{1-\alpha/2} ep(\hat{\beta}_j)\right) \approx 1-\alpha.$$

- Assim, um intervalo de confiança $1-\alpha$ (assintótico) para β_j fica dado por:

$$IC\left(\beta_j\,;1-\alpha\right)=\left(\hat{\beta}_j \pm z_{1-\alpha/2}\, ep\left(\hat{\beta}_j\right)\right).$$

No R: `confint.default(ajuste)`.

Intervalo de confiança para a resposta média em $\mathbf{x} = \mathbf{x_0}$

- A estimativa pontual da resposta média para um vetor de covariáveis $\mathbf{x}' = \mathbf{x_0}' = (1, x_{01}, x_{02}, \ldots, x_{0\,p-1})$, $\mu_0 = E[y \mid \mathbf{x_0}]$, baseada no ajuste de um modelo linear generalizado, é dada por:

$$\hat{\mu}_0 = g^{-1}\left(\mathbf{x_0}'\hat{\boldsymbol{\beta}}\right),$$

onde g é a função de ligação do modelo e $\hat{\boldsymbol{\beta}}$ a estimativa de máxima verossimilhança de $\boldsymbol{\beta}$.

- Seja $\hat{\eta}_0 = \mathbf{x_0}'\hat{\boldsymbol{\beta}}$ a estimativa do preditor linear calculada em $\mathbf{x_0}$. A variância assintótica de $\hat{\eta}_0$ fica dada por:

$$Var(\hat{\eta}_0) = Var\left(\mathbf{x_0}'\hat{\boldsymbol{\beta}}\right) = \mathbf{x_0}' Var\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x_0}.$$

- Como $\hat{\eta}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$ é uma combinação linear dos $\hat{\beta}'s$, temos que, assintoticamente:

$$\hat{\eta}_0 \sim Normal\left(\mathbf{x}_0'\boldsymbol{\beta}, \mathbf{x}_0' Var\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}_0\right).$$

- Assim, um intervalo de confiança $1-\alpha$ assintótico para $\eta_0 = \mathbf{x}_0'\boldsymbol{\beta}$ fica dado por:

$$IC(\eta_0, 1-\alpha) = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm z_{\alpha/2} \cdot \sqrt{\left(\mathbf{x}_0' V\hat{a}r\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}_0\right)},$$

sendo $z_{\alpha/2}$ o quantil $\alpha/2$ da distribuição Normal padrão. Apenas para efeito de notação, vamos representar o intervalo de confiança para η_0 por $\left(\eta_{0L}; \eta_{0U}\right)$.

- Assim, um intervalo de confiança assintótico $1-\alpha$ para μ_0 fica dado por:

$$IC(\mu_0, 1-\alpha) = \left(g^{-1}(\eta_{0L}); g^{-1}(\eta_{0U})\right),$$

se g for estritamente crescente e

$$IC(\mu_0, 1-\alpha) = \left(g^{-1}(\eta_{0U}); g^{-1}(\eta_{0L})\right)$$

se g for estritamente decrescente.

No R: `p1=predict(ajuste,type='link',newdata=x0,se.fit=T)`

```
### x0 é um dataframe com os dados para os quais se quer estimar a resposta.
### O argumento se.fit=T é para retornar os erros padrões das estimativas.
estimat=p1$fit
errpad=p1$se.fit
ic=exp(estimat+c(-1.96,1.96)*errpad) ### Vale se a ligação for logarítmica.
Se for outra, basta trocar exp() pela inversa da ligação usada.
```