Modelos Lineares II

PROFESSOR: RAFAEL ERBISTI

Modelo logístico geral

- O modelo de regressão logística mais geral assume

$$logit(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x}_i^T\boldsymbol{\beta}$$

onde $\boldsymbol{x}_i$ é o vetor de variáveis contínuas correspondendo a covariáveis ou variáveis *dummy*, e $\boldsymbol{\beta}$ é o vetor paramétrico.

- Esse modelo é amplamente usado para analisar dados envolvendo respostas binárias ou binomiais.

Modelo logístico geral

- Os dados podem ser agrupados como frequências para cada padrão de covariável (isto é, observações com os mesmos valores de todas as variáveis explicativas).

- Ou ainda, cada observação pode ser codificada 0 ou 1 e seu padrão de covariável ser listado separadamente.

- O processo de estimação é essencialmente o mesmo em ambos os casos.

Modelo logístico geral

- Se os dados podem ser agrupados, a resposta Y_i, o número de sucessos para o padrão de covariável i, pode ser modelado pela distribuição binomial.

- Se cada observação tem um padrão de covariável diferente, então $n_i = 1$ e a resposta Y_i é binária.

Exemplo: Suponha que testes clínicos sejam realizados para comparar a eficiência de um novo procedimento cirúrgico frente a uma técnica já conhecida. Os testes foram realizados em 2 hospitais ($x_1 = 1, 2$). Em cada hospital, os pacientes foram distribuídos aleatoriamente para serem submetidos a 2 procedimentos cirúrgicos ($x_2 = 1, 2$).

Modelo logístico geral

- No primeiro mês de estudo, sete pacientes foram recrutados. Estes pacientes são listados na tabela abaixo pelo seu número de identificação e pelas classes de covariáveis.

Dados listados pelo nº do paciente			Dados listados pela classe da covariável		
Paciente nº	Covariável(x_1, x_2)	Resposta(Y)	Covariável (x_1, x_2)	Tamanho(n)	Resposta(Y)
1	1,1	0	1,1	2	1
2	1,2	1	1,2	3	2
3	1,2	0	2,1	1	0
4	2,1	0	2,2	1	1
5	2,2	1			
6	1,2	1			
7	1,1	1			

Modelo logístico geral

- Os dados listados pelo padrão da covariável crescem em eficiência a medida que o número de pacientes aumenta.

- Nesse caso, as respostas tem a forma y_i/n_i onde $0 < y_i < n_i$ é o número de sucessos em n_i indivíduos no i-ésimo subgrupo (padrão, classe).

- Dados não agrupados podem ser considerados casos especiais em que $n_1 = n_2 = \cdots = n_N = 1$.

- O único problema em agrupar acontece se a ordem em que as observações aparecem for relevante.

Modelo logístico geral

- O estimador de máxima verossimilhança do parâmetro $\boldsymbol{\beta}$, e consequentemente das probabilidades $\pi_i = g^{-1}(\boldsymbol{x}_i^T\boldsymbol{\beta})$, são obtidas maximizando a função de log-verossimilhança

$$l(\boldsymbol{\pi};\boldsymbol{y}) = \sum_{i=1}^{N}\left[y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}\right]$$

usando o método escore iterativo.

- A deviance,

$$D = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right]$$

Modelo logístico geral

E que pode ser escrita como

$$D = 2 \sum o \log \frac{o}{e}$$

onde

o: frequências observadas y_i e $(n_i - y_i)$ das células da tabela de observações

e: frequências esperadas estimadas ou os valores ajustados $\hat{y}_i = n_i \hat{\pi}_i$ e $(n_i - \hat{y}_i) = (n_i - n_i \hat{\pi}_i)$.

- O somatório é realizado em todas as $2 \times N$ células da tabela de observações.

Modelo logístico geral

- Note que a deviance do modelo logístico não depende de parâmetro de ruído (σ^2 como no modelo normal).

- A bondade de ajuste pode ser avaliada e hipóteses podem ser testadas usando diretamente a aproximação

$$D \sim \chi^2(N-p)$$

onde p é o número de parâmetros estimados e N o número de padrões de covariável.

Modelo logístico geral

- Os métodos de estimação e as distribuições amostrais usadas para inferência dependem de resultados assintóticos.

- Para estudos pequenos ou situações onde existem poucas observações para cada padrão de covariável, os resultados assintóticos podem ser não muito adequados.

Outros critérios para verificar bondade de ajuste

A. Ao invés de usar o EMV poderíamos utilizar mínimos quadrados ponderados (estimando os parâmetros minimizando a soma ponderada dos quadrados)

$$S_w = \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1-\pi_i)}$$

já que $E(Y_i) = n_i\pi_i$ e $Var(Y_i) = n_i\pi_i(1-\pi_i)$.

- Isto é, equivalente a minimizar a estatística de qui-quadrado de Pearson

$$X^2 = \sum \frac{(o-e)^2}{e},$$

onde a soma é feita sob as $2N$ células da tabela de observações.

Outros critérios para verificar bondade de ajuste

- O motivo dessa equivalência é que:

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i} + \sum_{i=1}^{N} \frac{[(n_i - y_i) - n_i(1 - \pi_i)]^2}{n_i(1 - \pi_i)}$$

$$= \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1 - \pi_i)} (1 - \pi_i + \pi_i) = S_w$$

Outros critérios para verificar bondade de ajuste

- Quando X^2 é avaliada nas frequências estimadas, a estatística é dada por

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

que é **assintoticamente equivalente à deviance**

$$D = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - n_i\hat{\pi}_i}\right)\right].$$

Outros critérios para verificar bondade de ajuste

- A demonstração utiliza expansão em série de Taylor de $s \log(s/t)$ em torno de s, isto é:

$$s \log \frac{s}{t} = (s - t) + \frac{1}{2} \frac{(s-t)^2}{t} + \cdots$$

- Sob a hipótese de que o modelo é correto, tem-se, assintoticamente, que $D \sim \chi^2_{N-p}$ e, portanto, $X^2 \sim \chi^2_{N-p}$.

Outros critérios para verificar bondade de ajuste

- A escolha entre D e X^2 depende da adequabilidade da aproximação à distribuição $\chi^2(N-p)$.

- Há evidências que sugerem que, em geral, X^2 é melhor do que D, uma vez que D é influenciada por valores pequenos das frequências.

- Todas as aproximações tendem a ser pobres, se as frequências esperadas são muito pequenas.

Outros critérios para verificar bondade de ajuste

B. Algumas vezes utiliza-se a comparação entre função de log-verossimilhança do modelo ajustado e um modelo mínimo (nulo), para o qual $\pi_i = \pi, \forall i$.

- Sob o modelo nulo tem-se: $\tilde{\pi} = (\sum y_i)/(\sum n_i)$.

- Seja $\hat{\pi}_i$ a probabilidade estimada de Y_i sob o modelo de interesse $(\hat{y}_i = n_i \hat{\pi}_i)$.

Outros critérios para verificar bondade de ajuste

- A estatística é definida por

$$C = 2[l(\widehat{\boldsymbol{\pi}}; \boldsymbol{y}) - l(\widetilde{\boldsymbol{\pi}}, \boldsymbol{y})]$$

onde l é a função log-verossimilhança.

- Assim,

$$C = 2\sum\left[y_i \log\left(\frac{\hat{y}_i}{n_i\tilde{\pi}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - n_i\tilde{\pi}_i}\right)\right].$$

Outros critérios para verificar bondade de ajuste

- A distribuição amostral de C é, aproximadamente, $\chi^2(p-1)$ se todos os p parâmetros, exceto o intercepto, são iguais a zero (caso contrário, C terá uma distribuição não-central).

- Portanto, C é uma estatística de teste para a hipótese de que nenhuma das variáveis é necessária.

- C é chama de estatística de qui-quadrado da razão de verossimilhanças.

Outros critérios para verificar bondade de ajuste

C. Em analogia ao R^2 para o caso de regressão múltipla, outra estatística utilizada é

$$pseudo\ R^2 = \frac{l(\widetilde{\boldsymbol{\pi}}; \boldsymbol{y}) - l(\widehat{\boldsymbol{\pi}}; \boldsymbol{y})}{l(\widetilde{\boldsymbol{\pi}}; \boldsymbol{y})}$$

que representa ganhos proporcionais na função log-verossimilhança devido aos termos do modelo de interesse, quando comparado ao modelo nulo.

Outros critérios para verificar bondade de ajuste

D. O Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC) são outras estatísticas de qualidade de ajuste baseadas na função verossimilhança, com penalização para o número de parâmetros estimados e para a quantidade de dados:

$$AIC = -2l(\hat{\boldsymbol{\pi}}; \boldsymbol{y}) + 2p$$
$$BIC = -2l(\hat{\boldsymbol{\pi}}; \boldsymbol{y}) + p\ log(n)$$

onde p é o número de parâmetros estimados e n o total de observações.

Resíduos em regressão logística

- As questões estudadas na análise dos resíduos dos modelos de regressão múltipla para resposta contínua também são relevantes no contexto de respostas binárias:

  - Inclusão ou exclusão de covariáveis

  - Análise gráfica dos resíduos. Existem duas formas principais de resíduos correspondendo às medidas de bondade de ajuste D e X^2. Se existem m diferentes níveis de covariáveis, então podemos calcular m resíduos.

Resíduos em regressão logística

- Seja Y_k o número de sucessos, n_k o número de ensaios e $\hat{\pi}_k$ a probabilidade de sucesso estimada para o k-ésimo nível (padrão) de covariável.

- O resíduo de Pearson ou qui-quadrado é definido como

$$X_k = \frac{(y_k - n_k\hat{\pi}_k)}{\sqrt{n_k\hat{\pi}_k(1-\hat{\pi}_k)}}\,, \qquad\qquad k = 1, \dots, m$$

onde $\sum_{k=1}^{m} X_k^2 = X^2$, que é a estatística de qui-quadrado de Pearson da bondade de ajuste.

Resíduos em regressão logística

- Os resíduos de Pearson padronizados são

$$r_{Pk} = \frac{X_k}{\sqrt{1 - h_{kk}}}$$

onde h_{kk} mede o grau de influência da observação no ajuste do modelo (leverage), e é obtido por meio do k-ésimo elemento da diagonal da matriz $\boldsymbol{H}$.

Resíduos em regressão logística

- Um segundo tipo de resíduo é o resíduo da deviance

- O valor total da deviance pode ser escrito na forma $D = \sum_{k=1}^{m} d_k^2$ em que cada componente individual é dada por

$$d_k = sign(y_k - n_k\hat{\pi}_k)\left\{2\left[y_k \log\left(\frac{y_k}{n_k\hat{\pi}_k}\right) + (n_k - y_k)\log\left(\frac{n_k - y_k}{n_k - n_k\hat{\pi}_k}\right)\right]\right\}^{1/2}$$

onde o termo $sign(y_k - n_k\hat{\pi}_k)$ garante que d_k tenha o mesmo sinal de X_k.

Resíduos em regressão logística

- Os resíduos padronizados baseados na deviance são definidos por

$$r_{Dk} = \frac{d_k}{\sqrt{1 - h_{kk}}}.$$

- As análises de resíduos em MLG devem ser conduzidas da mesma maneira que em modelos lineares normais.
- Se os dados são binários, ou se n_i é pequeno para a maioria dos níveis das covariáveis, então haverá poucos valores distintos dos resíduos e os gráficos serão pouco informativos.
- Neste caso, será necessário confiar na estatística agregada da bondade de ajuste X^2 e D, e outros diagnósticos.

Sobredispersão

- Sobredispersão ou variação extra-binomial é um fenômeno comum que ocorre na modelagem de dados binários agrupados.

- Sua ocorrência é caracterizada quando a variação observada excede aquela assumida pelo modelo.

- Ou seja, observações Y_i que são assumidas seguir uma distribuição binomial podem apresentar variância maior do que $n_i\pi_i(1-\pi_i)$.

Sobredispersão

- Uma abordagem é incluir um parâmetro extra, ϕ, de modo que

$$Var(Y_i) = \phi\, n_i \pi_i (1 - \pi_i)$$

  - Se $\phi = 1$ temos variabilidade binomial.
  - Se $\phi > 1$ temos extra variabilidade.

- Pode haver indícios de sobredispersão quando as estatísticas de qualidade de ajuste (deviance e X^2 de Pearson) são grandes em relação aos seus graus de liberdade, $(N - p)$.

Sobredispersão

- Entretanto, deve-se lembrar que alguns pontos aberrantes podem aumentar substancialmente o valor do desvio e a “simples eliminação desses pontos pode reduzir as evidências de sobredispersão”.

- Para investigar o efeito de observações influentes as estatísticas delta-beta, delta-qui-quadrado e delta-deviance também estão disponíveis para regressão logística.

- A sobredispersão tem duas causas possíveis:
  - um modelo especificado incorretamente: seriam necessários no modelo mais termos, tais como interações ou termos quadráticos;
  - falta de independência das observações.

Sobredispersão

- O parâmetro ϕ pode ser estimado com base na estatística de Pearson ou na deviance, respectivamente por:

$$\tilde{\phi} = \frac{X^2}{N-p} = \frac{1}{N-p}\sum_{i=1}^{N}\left(\frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}}\right)^2$$

$$\tilde{\phi} = \frac{D}{N-p} = \frac{1}{N-p}\sum_{i=1}^{N} d_i^2$$

onde $N-p$ é o total de graus de liberdade.

- Na prática, multiplicamos a raiz quadrada de $\tilde{\phi}$ pelos desvios-padrão estimados dos β's para construir intervalos de confiança e fazer testes de hipóteses.

Conceito de chance

- Uma forma natural de quantificar as chances de um evento é utilizando probabilidades.

- Outra forma de fazer isso é a partir da razão de probabilidades.

- Se A e B são eventos tais que $A \cap B = \emptyset$ e $A \cup B = \Omega$, a razão de probabilidades

$$\frac{P(A)}{P(B)} = \frac{P(A)}{1 - P(A)}$$

é denominada de chances (odds) do evento A relativo ao evento B.

Conceito de chance

- As chances do evento A também pode ser calculada como a razão entre o número de vezes que A ocorre sobre o número de vezes que ele não ocorre.

Exemplo: Uma chance de 4 significa que esperamos que ocorrências sejam 4 vezes as não ocorrências do evento.

Exemplo: A probabilidade de nascimento de um indivíduo do sexo masculino é cerca de 0,515. Então a chance desse evento é 0,515/0,485=1,062. A chance em favor do nascimento de um indivíduo do sexo masculino é 106 para 100 ou 106 nascimentos masculinos para 100 femininos.

- O modelo logístico pressupõe que o logaritmo da chance é linearmente relacionado com as variáveis explicativas.

Razão de chances

- Considere inicialmente o modelo logístico linear simples em que $\pi(x)$, a probabilidade de "sucesso" dado o valor x de uma variável explicativa qualquer, é definida tal que

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x,$$

em que β_0 e β_1 são parâmetros desconhecidos.

- Esse modelo poderia, por exemplo, ser aplicado para analisar a associação entre uma determinada doença e a ocorrência ou não de um fator particular.

Razão de chances

- Seriam então amostrados, independentemente, n_1 indivíduos com presença do fator $(x = 1)$ e n_2 indivíduos com ausência do fator $(x = 0)$ e $\pi(x)$ seria a probabilidade de desenvolvimento da doença após um certo período fixo.

- Dessa forma, a chance de desenvolvimento da doença para um indivíduo com presença do fator fica dada por

$$\frac{\pi(1)}{1-\pi(1)} = \exp(\beta_0 + \beta_1)\,,$$

enquanto que a chance de desenvolvimento da doença para um indivíduo com ausência do fator é simplesmente

$$\frac{\pi(0)}{1-\pi(0)} = \exp(\beta_0)\,.$$

Razão de chances

- Logo, a razão de chances fica dada por

$$\psi = \frac{\pi(1) \big/ \left(1-\pi(1)\right)}{\pi(0) \big/ \left(1-\pi(0)\right)} = \frac{\pi(1)\{1-\pi(0)\}}{\pi(0)\{1-\pi(1)\}} = \exp(\beta_1)$$

dependendo apenas do parâmetro β_1.

- Uma das grandes vantagens da regressão logística é a possibilidade de interpretação direta dos coeficientes como medidas de associação.

Razão de chances

Exemplo: Suponha que a variável resposta corresponda ao uso de anticoncepcional e estamos interessados na razão do número esperado de usuários para cada não usuário (chance). Como variável explicativa temos um fator com dois níveis: urbano (1) e rural (0). Suponha que as chances em favor do uso sejam de 4 para 1 em áreas urbanas e de 2 para 1 em áreas rurais.

- Então a razão de chances nas áreas urbanas para as chances em áreas rurais é 2.

- Neste caso, o número esperado de usuários para cada não usuário em áreas urbanas é duas vezes que em áreas rurais.

- Em termos de porcentagem: o número esperado de usuários para cada não usuário é 100% maior em áreas urbanas comparado às áreas rurais.

Razão de chances

- O modelo nesse caso é:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x\ ,$$

onde $x = 1$, se urbana e $x = 0$, se rural.

- Para $x = 1 \ \Rightarrow \ \frac{\pi_1}{1-\pi_1} = \exp(\hat{\beta}_0 + \hat{\beta}_1) = 4$
- Para $x = 0 \ \Rightarrow \ \frac{\pi_0}{1-\pi_0} = \exp(\hat{\beta}_0) = 2$
- Então $\left(\frac{\pi_1}{1-\pi_1}\right)/\left(\frac{\pi_0}{1-\pi_0}\right) = \exp(\hat{\beta}_1) = \frac{4}{2} = 2$ (a chance em áreas urbanas é 2 vezes a chance em áreas rurais quando comparamos usuários e não usuários).

Razão de chances

- No caso em que a razão de chances resulta em um número menor que 1 a interpretação pode ser feita da seguinte forma:

- Suponha que no exemplo a razão de chances resulte em 0.2, então, a chance em áreas rurais é 5 vezes a chance em áreas urbanas quando comparamos usuários e não usuários de anticoncepcional.

- Ou ainda, em termos de porcentagem, o número esperado de usuários para cada não usuário é 80% menor em áreas urbanas comparado às áreas rurais. (a chance em áreas urbanas é 80% menor do que em áreas rurais).

Razão de chances

Exemplo: Estudo para avaliar o efeito de taxa e volume de ar que uma pessoa inspira na probabilidade de ocorrência de um acidente vascular.

Resposta igual a 1 - sucesso, ocorrência do evento

Resposta igual a 0 - fracasso, não ocorrência do evento

- Neste caso, os dados aparecem não agrupados e as variáveis explicativas foram medidas em escala contínua.

Razão de chances

- O modelo a ser ajustado é da forma:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 taxa_i + \beta_2 volume_i$$

onde π_i é a probabilidade de que o i-ésimo indivíduo sofra um acidente vascular.

Razão de chances

```
> summary(modelo1)

Call: glm(formula = resposta ~ taxa + volume, family
=binomial(link = "logit"),    data = resp)

Coefficients:
              Estimate    Std. Error    z value   Pr(>|z|)
(Intercept)    -9.5296        3.2332     -2.947   0.00320 **
taxa            2.6491        0.9142      2.898   0.00376 **
volume          3.8822        1.4286      2.717   0.00658 **
---
(Dispersion parameter for binomial family taken to be 1)

   Null deviance: 54.040  on 38  degrees of freedom
Residual deviance: 29.772  on 36  degrees of freedom AIC: 35.772

Number of Fisher Scoring iterations: 6
```

Razão de chances

```
> anova(modelo1)
Analysis of Deviance Table

Model: binomial, link: logit

Response: resposta

Terms added sequentially (first to last)

       Df  Deviance Resid.  Df  Resid. Dev
NULL                        38      54.040
taxa    1        4.385      37      49.655
volume  1       19.883      36      29.772
```

Razão de chances

- A chance estimada do i-ésimo indivíduo sofrer um acidente vascular é dada por

$$\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 taxa_i + \hat{\beta}_2 volume_i\right)$$

- A chance estimada do terceiro indivíduo sofrer um acidente vascular é

$$\left(\frac{\hat{\pi}_3}{1-\hat{\pi}_3}\right) = \exp\{-9{,}530+2{,}649(2{,}5)+3{,}882(1{,}25)\} = 6{,}994$$

- Assim, para uma pessoa que tem taxa de ar inspirado igual a 2,5 e volume igual 1,25 a ocorrência de um acidente vascular é de aproximadamente 7 para 1.

Razão de chances

- Vamos observar o que ocorre quando variamos a taxa em uma unidade e mantemos o volume de ar constante:

$$\left(\frac{\hat{\pi}_i/1-\hat{\pi}_i}{\hat{\pi}_j/1-\hat{\pi}_j}\right)=\frac{\exp\left\{\hat{\beta}_0+\hat{\beta}_1(taxa+1\ )+\hat{\beta}_2 volume_i\right\}}{\exp\left\{\hat{\beta}_0+\hat{\beta}_1 taxa+\hat{\beta}_2 volume_i\right\}}=\exp(\hat{\beta}_1)$$

- Então, para cada unidade acrescida na taxa de ar inspirado, mantendo-se o volume constante, a razão de chances aumenta em

$$\exp\left(\hat{\beta}_1\right)=\exp(2{,}649)=14{,}14$$

Razão de chances

- De forma análoga, para um aumento de uma unidade do volume de ar inspirado, mantendo-se a taxa constante, a razão de chances aumenta em

$$\exp\left(\hat{\beta}_2\right) = \exp(3{,}882) = 48{,}52.$$

- Como ambos coeficientes estimados são positivos, tem-se que aumentos nas variáveis, implicam em aumentos na chance de ocorrência do evento.

- Verifique que, pela simetria da função logística, se tivéssemos modelado o evento complementar não ocorrência do acidente vascular, obteríamos os coeficientes com sinais trocados, mas a estatística Deviance permaneceria igual.

Intervalo de confiança para a razão de chances

- Considere o modelo

$$Y_i \sim \; Bin(n_i, \pi_i)$$

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x$$

onde x é um fator em que se $x = 0$, ausência do fator, se $x = 1$ presença do fator.

- Usualmente, é mais fácil interpretar os efeitos das variáveis explicativas em termos das razões de chances do que olhar para os parâmetros β.

Intervalo de confiança para a razão de chances

- Baseado no modelo, podemos obter o quanto as chances aumentam na presença do fator. Então as chances são:

$$\frac{\pi(0)}{1-\pi(0)} = \exp(\beta_0)$$

quando $x = 0$ indicando que o fator está ausente, e

$$\frac{\pi(1)}{1-\pi(1)} = \exp(\beta_0 + \beta_1)$$

quando $x = 1$ indicando que o fator está presente.

Intervalo de confiança para a razão de chances

- Portanto,

$$\psi = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} = \exp(\beta_1)$$

- Se $\beta_1 = 0$ então $\psi = 1$, o que corresponde um "não efeito" da presença do fator.

- Denotando a razão de chances por ψ_j, no modelo de regressão logística, temos

$$\psi_j = \exp\left(\beta_j\right)$$

e podemos estimá-lo por $\exp(\hat{\beta}_j)$.

Intervalo de confiança para a razão de chances

- Assim o intervalo de confiança assintótico para ψ_j com nível de confiança $100(1-\alpha)\%$ terá limites

$$\exp\left(\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\beta}_j)}\right)$$

- Por exemplo, intervalos de 95% de confiança para ψ_j são calculados através de

$$\exp\left(\hat{\beta}_j \pm 1{,}96\sqrt{Var(\hat{\beta}_j)}\right)$$

- Intervalos que não incluem a unidade correspondem a valores de β significativamente diferentes de zero.