Disciplina:

Técnicas de Amostragem e Modelos de Regressão

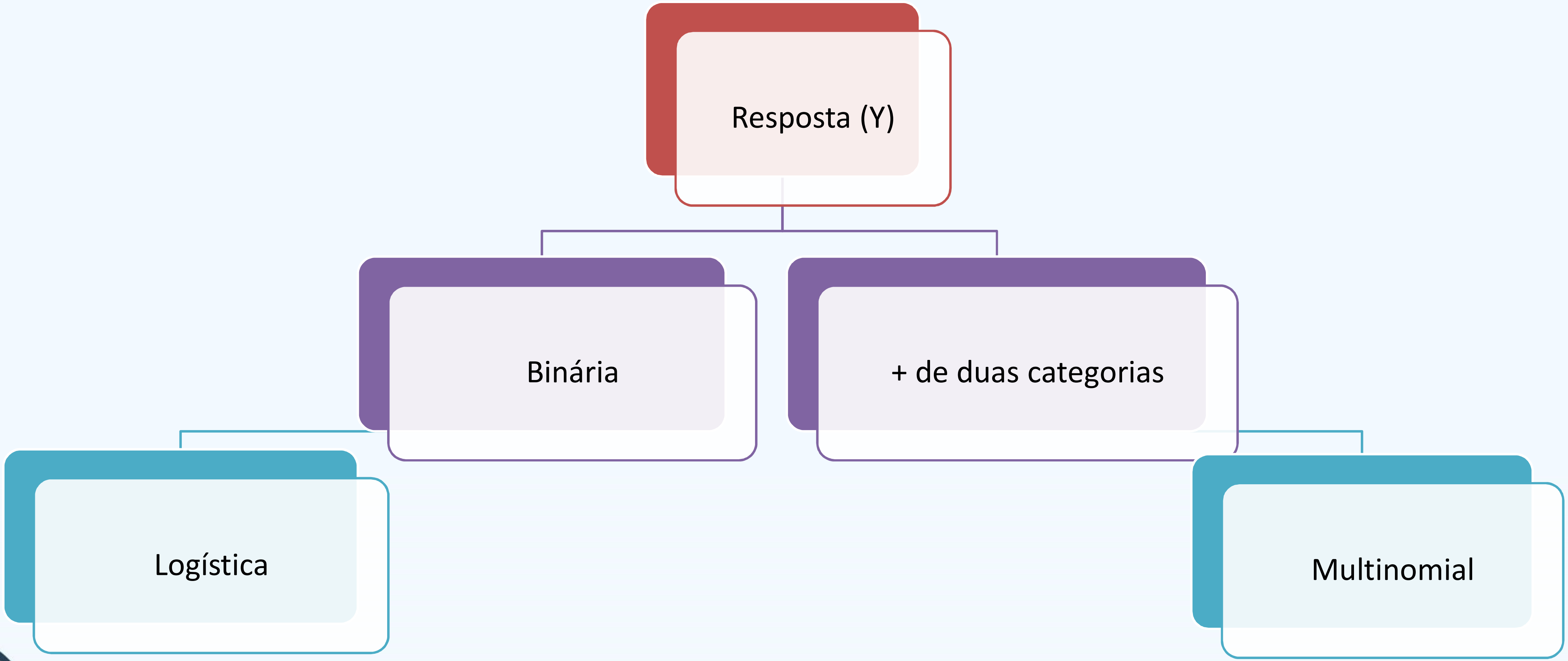
Professor: Anaíle Mendes Rabelo

PUC Minas

Regressão Multinomial

PUC Minas

GLM – Regressão Binária



PUC Minas

Recapitulando a Regressão Logística

Suponha que o modelo tenha a seguinte forma:

$$\eta = X'B$$

Em que $X' = [1, X_{i1}, X_{i2}, \ldots, X_{in}]$, $\mathrm{B} = [\beta_0, \beta_1, \beta_2, \ldots, \beta_n]$, e a variável resposta entre 0 e 1.

Assumimos que a variável reposta é uma variável aleatória Binomial.

Recapitulando a Regressão Logística

Probabilidade de sucesso:

$$p = \frac{nº\ de\ sucesso}{nº\ de\ tentativas}$$

Probabilidade de falha :

$$q = \frac{nº\ de\ falhas}{nº\ de\ tentativas} \qquad q = 1 - p$$

Chance de sucesso (Odds):

$$\text{Odds} = \frac{nº\ de\ sucesso}{nº\ de\ falhas} = \frac{p}{1-p}$$

PUC Minas

Regressão logística / Binomial binária

Seja Y uma variável aleatória binária com distribuição binomial de probabilidade de sucesso π(x).

A notação π(x) sugere que a probabilidade de sucesso está condicionada a um valor/categoria x.

Ou seja, π(x)=Pr(Y=1|X=x). Define-se então:

$$\eta = x'\beta$$

$$\eta = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$$

Regressão logística / Binomial binária

- Na regressão binomial binária ou regressão logística clássica temos que Y poderia tomar dois valores:

  – $Pr(Y=0|X=x)$.
    $Pr(Y=1|X=x)$.

Regressão Multinominal

- Seja Y uma variável aleatória categórica com J categorias.
- Seja $\pi_j(x) = \Pr(Y = j \mid x), com\ \sum_j \pi_j(x) = 1$
- O modelo compara cada categoria j com uma categoria de referência J, totalizando $\binom{J}{2}$ combinações.

$$\eta = \log\left(\frac{\pi_j(x)}{\pi_k(x)}\right) = \alpha_j + \beta'_j x = \frac{1}{1+e^{-(\alpha_j+\beta_j x)}}$$

SEMPRE IREMOS OLHAR EM RELAÇÃO A UMA CATEGORIA DE REFERÊNCIA

PUC Minas

Regressão Multinominal / Binomial

- Variável resposta: Qualquer nº de níveis

- Ex: Estudo da escolha de um plano de saúde.
  - Y : Tipos de plano (A,B,C)
  - X: Idade, tamanho da família, renda, etc.

- A regressão logística Multinominal pode também ser chamada de Regressão Logística Politômica

PUC Minas

Regressão Multinominal / Binomial

Considerando Y com 3 categorias:

Modelo necessita de 2 funções.

Comparação de categorias:

- Y =0 -> Referencia
- Comparar com Y =1 e Y =2.

$$g_1(x) = \ln\left(\frac{P(Y=1\,)|\,x}{P(Y=0\,)|\,x}\right) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p = x'\beta_1$$

$$g_2(x) = \ln\left(\frac{P(Y=2\,)|\,x}{P(Y=0\,)|\,x}\right) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p = x'\beta_1$$

PUC Minas

Interpretação dos Parâmetros Regressão Multinomial

PUC Minas

Regressão Multinomial

- Seja Y uma variável aleatória categórica com J categorias.
- Seja $\pi_j(x) = \Pr(Y = j \mid x), com \sum_j \pi_j(x) = 1$
- O modelo compara cada categoria j com uma categoria de referência J, totalizando $\binom{J}{2}$ combinações.

$$\eta = \log\left(\frac{\pi_j(x)}{\pi_k(x)}\right) = \alpha_j + \beta'_j x = \frac{1}{1+e^{-(\alpha_j+\beta'_j x)}}$$

SEMPRE IREMOS OLHAR EM RELAÇÃO A UMA CATEGORIA DE REFERÊNCIA

PUC Minas

Regressão Multinominal / Binomial

Considerando Y com 3 categorias:

Modelo necessita de 2 funções.

Comparação de categorias:

- Y =0 -> Referencia
- Comparar com Y =1 e Y =2.

$$g_1(x) = \ln\left(\frac{P(Y=1)|\,x}{P(Y=0)|\,x}\right) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p = x'\beta_1$$

$$g_2(x) = \ln\left(\frac{P(Y=2)|\,x}{P(Y=0)|\,x}\right) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p = x'\beta_1$$

PUC Minas

Regressão Multinominal

Resposta (Y) entre 0 e 1

Requer transformação da nossa função

Função Logit

$$E(y) = \frac{e^{x\prime\beta}}{1 + e^{x\prime\beta}} = \frac{1}{1 + e^{-x\prime\beta}}$$

Regressão Multinominal

A regressão logística pode ser linearizada:

$$\eta = x'\beta$$

Ser o preditor linear, onde η é definido pela transformação logit.

$$\eta = ln \frac{p}{1 - p}$$

A razão $\frac{p}{1 - p}$ é chamada de chance odds .

PUC Minas

Estimação de Parâmetros

- A estimação dos parâmetros de $x'_i\beta$ é realizada a partir do método de máxima verossimilhança;
- Como nosso dados seguem a distribuição de Binomial, então a distribuição de probabilidade é dada por:

$$f_i\ (y_i; n; p) = \binom{n}{y} p^y (1 - p)^{n-y}$$

- Logo a função de verossimilhança para v.a. independentes pode ser dada por:

$$\mathrm{L}(y_1, y_2, y_3, \dots, y_n, \beta) = \prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \binom{n}{y} p^y (1 - p)^{n-y}$$

Interpretação dos Parâmetros – ODDS RATIO

Razão entre a probabilidade de um evento ocorrer (p(Y=1)) e a probabilidade de não ocorrer (p(Y=0)).

$$\frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} = \frac{e^{\hat{\beta}_0+\hat{\beta}_1}}{e^{\hat{\beta}_0}} = e^{\hat{\beta}_1}$$

Chance de sucesso em relação à chance de fracasso.

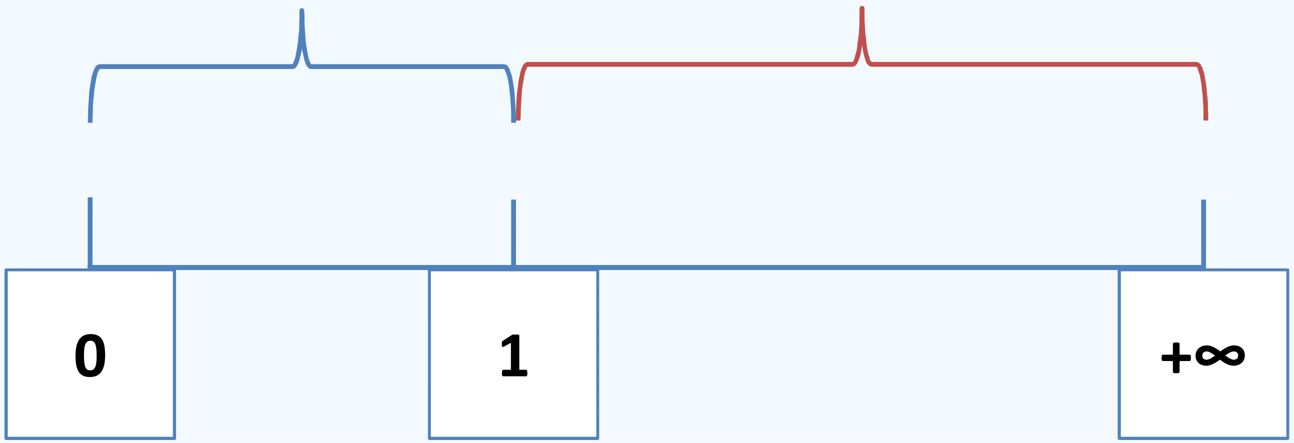
$$g_1(x) = \ln\left(\frac{P(Y=1\,)|\,x}{P(Y=0\,)|\,x}\right) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p = x'\beta_1$$

$$g_2(x) = \ln\left(\frac{P(Y=2\,)|\,x}{P(Y=0\,)|\,x}\right) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p = x'\beta_1$$

Interpretação dos Parâmetros – ODDS RATIO

$$\widehat{O_R} = \frac{odds_{x_i+1}}{odds_{x_i}} = e^{\widehat{\beta}_1}$$

Reduz a probabilidade de ocorrência | Aumenta a probabilidade de ocorrência



Pressupostos da Regressão Multinominal

PUC Minas

- A variável reposta precisa ser qualitativa, com mais de duas categorias
- As preditoras podem ser quantitativas ou categóricas
- Ausência de autocorrelação
- Relação linear entre o vetor das variáveis explicativas X e a variável independente Y;
- Ausência de correlação entre os resíduos
- Assume que as observações são independentes
- Ausência de multicolinearidade

Verificar a significância das variáveis do modelo

Teste de hipótese para determinar se a variável preditora do modelo é significativamente relacionada com variável resposta do modelo

- Teste de Wald
- Teste de Razão de verossimilhança

Análise dos resíduos

- Resíduos de Pearson
- Resíduos de Deviance
- Resíduos de Pseudo –Valor
- Gráfico de Resíduos