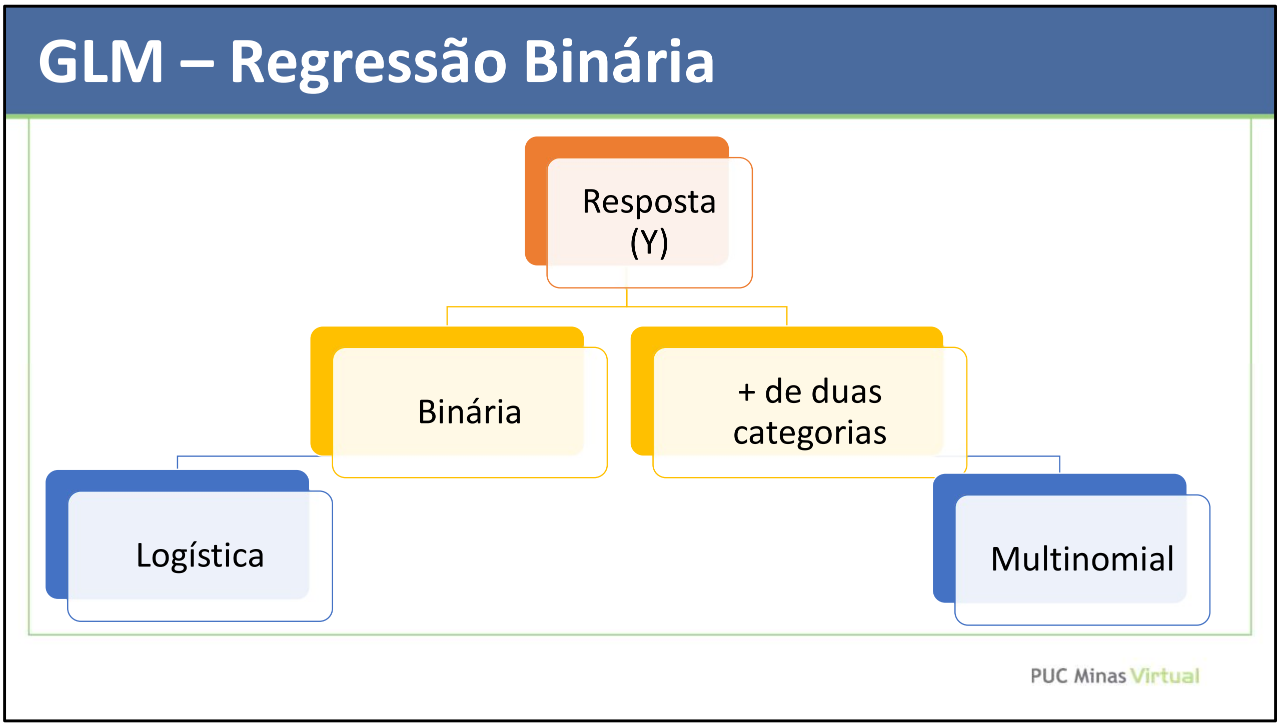
Regressão Multinomial

Nesta aula iremos entender sobre a interpretação dos parâmetros de uma regressão multinomial



Então apenas um comparativo com os modelos anteriormente estudados.
Na regressão logística temos apenas duas possíveis respostas. Sucesso ou fracasso, ou 0 e 1

Ou seja nossa variável agora, é uma variável categórica, podemos lembrar aqui das variáveis dummy , onde tínhamos uma variável qualitativa com duas categorias e criamos uma variável 0 e1 para conseguirmos trabalhar numericamente com elas. Então é exatamente o que ocorre na regressão logística.

Porem tmabem podemos ter mais de 2 categorias em nossa variável repsostae quando isso ocorre estamos falando no caso de uma regressão chamada multinomial. Que também sera um modelo linear generalizado

Recapitulando a Regressão Logística

Suponha que o modelo tenha a seguinte forma:

$$\eta = X'B$$

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

Assumimos que a variável reposta é uma variável aleatória Binomial.

PUC Minas Virtual

Então para recapitular tudo que foi criado em regressão logística, pois iremos observar que muitas coisas serão reaproveitadas.

Entao temos o modelo com nosso componente sistemático XB e nossa matriz de erros aleatórios (componente aleatório)

Observe que nesse modelo como já vimos, não estou separando a variável logística com uma ou várias variáveis preditivas, para o cálculo dos estimadores, vimos que isso não irá importar utilizando o conceito matricial.

Na forma tradicional da regressão linear teríamos um problema no Y, pois iria ultrapassar 0 e 1

Como a nossa variável resposta agora é uma probabilidade, inicialmente, com duas possibilidade, então iremos assumir que ela segue uma distribuição de bernoulli.

Na distribuição de bernoulli , estudasse o sucesso e a probabilidade de se obter o sucesso.

Como a variável resposta segue uma distribuição de bernoulli .

Temos que Y = 1 seria a prob de sucesso que pode ser descrita por pi

E no caso contrário, Y = 0, ou fracasso terimos entao a probabilidade complementar do sucesso , que é dada por 1 - p

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 Virtual

Entao podemos pontuar alguns conceitos importantes desse modelo

A probabilidade de sucesso , ous eja de obtermos y = 1 , pode ser definida com o número de sucesso sobre o número de tentativas .

Já a probabilidade de fracasso , ous eja de obtermos y = 0 , pode ser definida com o número de falhas sobre o número de tentativas .

O que foi utilizado e calculado anteriormente como 1 – p

E a odds ou chance pode ser definida como a probabilidade de sucesso sobre a probabilidade de fracasso.

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:

$$\eta = x'\beta$$

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

PUC Minas Virtual

Logo , vimos que a regressão binaria, ou logística , é aplicável quando temos que Y é uma variável aleatória binaria com probabilidade de sucesso pi(x) ou p

E observe que a probabilidade de sucesso sempre estará condicionada a um valor ou categoria de x

Ou seja , pi(x) que é a probabilidade de Y ser 1 dado que x = x é dado por

Eta = ao componente sistemático que pode ser dado por x transpoto beta

Que sera o log (p/ 1 – p) = b0 + b1x =

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

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).

PUC Minas Virtual

Observe que na regressão logística temos que Y pode assumir apenas 2 valores 0 e 1, então teríamos duas hipóteses de estudo de probabilidade

Probabilidade de y = 0 dado o valor de x

E a probabilidade de y = 1 dado um certo valor 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.

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

SEMPRE IREMOS OLHAR EM RELAÇÃO A UMA CATEGORIA DE REFERÊNCIA

PUC Minas Virtual

Agora já pensando em uma regressão multinomial

Temos que nossa variável resposta pode assumir ate J categorias.

Logo agora iremos pensar na probabilidade de sucesso de Y ser igual a j dado um valor de x, observe que pela teoria da probabilidade o somatório dessas probabilidade continua sedo igual a 1.

Entao para o modelo multinomial vamos ter o seguinte,

Iremos sempre olhar em relação a uma categoria diferente , ou seja iremos sempre comparar as categorias de dois a dois.

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 Virtual

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

Tao temos que como variável resposta agora poderemos ter qualquer quantidade de níveis ou categorias

Entao por exemplo podemos comparar 3 tipos de planos de saúde e verificar qual a probabilidade da pessoa escolher o tipo A b e c dados a idade , tamenho da família e renda por exemplo

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 Virtual

E para isso iriamos considerar a comparação das categorias de 2 em duas. Logo iriamos ter a mesma ideia da regressão pois transformaríamos as 3 variáveis em comparações de dois a dois. E sempre teríamos uma determinada categoria de referência. Neste caso a categoria 0.

PUC Minas
Virtual