Aula 11: Modelos não lineares

**Lista de Exercícios**

30/11 & 02/12/2020

Esta lista foi adaptada a partir de <https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/> e o código completo para a realização da atividade está em <https://rpubs.com/bezerra_mila/logit>

**Como interpretar Razões de Chance (*Odds ratio*)?**

Devemos partir da ideia de probabilidade. Se um evento possui 75% de chances de ocorrer, há, obviamente, 25% de chances de não ocorrer. As chances de sucesso são definidas como a razão da probabilidade de sucesso pela probabilidade de fracasso. Neste exemplo, as chances de sucesso são de 75%/25% = 3. Isto significa dizer que as chances de sucesso são de 3 para 1. Esse valor, geralmente, é expresso em sua forma logarítmica. Neste caso, *ln (3) = 1,099*. Esta conversão para logaritmo é feita porque as probabilidades estão sempre restritas a um único intervalo e a transformação tem o intuito de romper estes limites. A segunda razão é que esta é a transformação logística, que é uma forma fácil de interpretar os resultados.

Sabemos pelo modelo de regressão logística que a sua função é expressa por:

Podemos interpretar *Y* como a probabilidade *p* de sucesso de um evento qualquer. Esta probabilidade é explicada pelos fatores *X* de acordo com a função logística. Vamos fazer algumas transformações. Para tornar mais simples, vamos substituir *Y* por *p* e vamos inverter os termos:

Esta expressão pode ser reescrita separando os termos do lado direito da equação:

Podemos, então, subtrair 1 dos dois lados da equação e rearranjar os termos:

Nesta expressão, temos do lado esquerdo, a probabilidade de fracasso dividida pela probabilidade de sucesso. Esta é exatamente o inverso da razão de chance expressa anteriormente. Além disso, do lado direito, não só temos o inverso da função de associação de *y* com *x*, como ela ainda permanece como um expoente. A fim de encontrarmos a função que nos interessa, são necessárias duas passagens: a primeira é a inversão das frações e a segunda, a extração do expoente. Para esta última, usamos o logaritmo:

Esta expressão da esquerda pode ser reescrita como a função logística de *p*, e daí, teríamos:

Note, então, que a partir da razão de chances, temos uma expressão linear em *x*. A razão de chance é, então, uma alteração na variável dependente, afetando, assim, a maneira como interpretamos os coeficientes estimados.

Vamos, então, lidar com uma base de dados para entender melhor o que as razões de chance significam e como elas são apresentadas como resultado de um modelo *logit*.

**Exemplo prático**

Repita estes códigos ao longo do exercício e interprete os resultados.

Vamos começar instalando alguns pacotes e acessando a base da UCLA:

*#Pacotes utilizados*

pacotes <- c("rgl","car", "tidyverse",

 "reshape2","jtools","lmtest","caret","pROC","ROCR","nnet")

**if**(sum(as.numeric(!pacotes %**in**% installed.packages())) != 0){

 instalador <- pacotes[!pacotes %**in**% installed.packages()]

 **for**(i **in** 1:length(instalador)) {

 install.packages(instalador, dependencies = T)

 **break**()}

 sapply(pacotes, **require**, character = T)

} **else** {

 sapply(pacotes, **require**, character = T)

}

link <- 'https://stats.idre.ucla.edu/wp-content/uploads/2016/02/sample.csv'

ucla <- read.csv(link)

glimpse(ucla)

Esta base de dados possui 200 observações e a variável resposta é chamado ***hon***, indicando se um estudante está em uma “*honor class”* ou não. Neste caso, então, .

Vamos começar, então, considerando um modelo sem nenhuma variável explicativa. Neste caso, temos:

modelo1 <- glm(hon ~1,

 data = ucla,

 family = "binomial")

*#visualizando os resultados*

summ(modelo1, confint = T, digits = 3, ci.width = .95)

Como encontrar, a partir deste resultado, a probabilidade *p*?

Sabemos aqui que . E também sabemos que *p* é a chance geral de um aluno estar em uma “*honor class”*.

Vamos, então, considerar a frequência absoluta de *hon*:

table(ucla$hon)

E também sua frequência relativa:

table(ucla$hon)/nrow(ucla)\*100

1. Determine a probabilidade, a razão de chance e seu logaritmo.

Em outras palavras, o intercepto do modelo sem variável explicativa é o logaritmo das chances estimado de estar em uma “*honor class”* para qualquer indivíduo da população de interesse.

1. O que este resultado significa?

Vamos adicionar agora uma variável explicativa, uma *dummy* se o indivíduo é uma mulher (=1) ou não:

modelo2 <- glm(hon ~ female,

 data = ucla,

 family = "binomial")

*#visualizando os resultados*

summ(modelo2, confint = T, digits = 3, ci.width = .95)

Antes de interpretarmos os resultados, vamos considerar a frequência conjunta das duas variáveis:

xtabs(~hon + female, data = ucla)

1. Encontre a probabilidade de um homem estar na “*honor class”,* a probabilidade de uma mulher estar na “*honor class”* e qual é a razão de chance para as mulheres em relação aos homes.

Vamos agora incluir uma variável explicativa contínua, que é a nota obtida nas provas de matemática. Nossa equação agora fica:

Vamos encontrar as estimativas do modelo.

modelo3 <- glm(hon ~ math,

 data = ucla,

 family = "binomial")

summ(modelo3, confint = T, digits = 3, ci.width = .95)

1. Como interpretamos a constante deste modelo? Este valor é observado na amostra selecionada ou é uma extrapolação? Como isto altera a nossa interpretação?

Para interpretarmos o valor do coeficiente da variável *math*, precisamos considerar a *variação* de seu valor para avaliarmos a mudança no logaritmo da razão de chance.

1. Tome um valor qualquer possível para a variável *math*, encontre o logit para este valor de *x*, repita o processo para *x + 1* e determine a variação do logaritmo da razão de chance.
2. Como encontrar o valor da mudança das chances para uma variação em 1 ponto em matemática? Esta mudança depende do valor de matemática que se obteve?
3. Rode um modelo agora com as variáveis *math* e *female* em conjunto. Interprete os resultados.
4. O que aconteceria se ao invés de rodarmos um modelo não linear, utilizássemos um comando de regressão de OLS convencional?
	1. Estime os parâmetros via OLS e interprete os resultados
	2. Compare com os resultados obtidos no item 7 acima.

Sugiro a todos que façam os exercícios para a seção em que a interação é discutida.