

Exemplo aula 03_Atividade_prática 1.2

Cynthia Tojeiro

2025-02-12

#Modelos de Regressão Múltipla Inicialmente vamos carregar os dados no R, utilizando os códigos a seguir.

```
# Limpa a memória
```

```
#rm(list=ls())
```

```
# Pacote necessario para leitura dos dados
```

```
library(readxl) #para ler no excell
```

```
# Dados referentes a diabetes mellitus tipo 1 (DM1)
```

```
dados <- read.table("C:\\datascience\\exercicio1.txt")
```

```
#Para ver as 6 primeiras linhas do conjunto de dados
```

```
head(dados)
```

```
##      V1 V2 V3 V4      V5  V6 V7      V8
## 1   1  1  57 11 25.8  56  0 7.413015
## 2   2  2  67 17 29.6 189  0 6.252037
## 3   3  3  42  9 25.2 122  1 4.335380
## 4   4  4  0 62  8 24.6 169  1 7.390550
## 5   5  5  1 50  8 20.2 133  0 5.463982
## 6   6  6  1 62  8 26.2 172  0 7.458546
```

```
attach(dados)
```

```
sexo <- dados[,2]
```

```
idade<-dados[,3]
```

```
escolaridade<-dados[,4]
```

```
imc<-dados[,5]
```

```
tempodiabetes<-dados[,6]
```

```
usoinsulina<-dados[,7]
```

```
hemoglobina<-dados[,8]
```

#Modelos de Regressão Múltipla Vamos continuar com os mesmos dados, entretanto agora para fazer o ajuste do modelo de regressão múltipla usamos a função lm da seguinte forma:

```
fit.model<-result<-lm(hemoglobina~sexo+idade+escolaridade+imc+tempodiabetes+usoinsulina)
```

```
summary(result)
```

```
##
```

```
## Call:
```

```
## lm(formula = hemoglobina ~ sexo + idade + escolaridade + imc +
```

```
##      tempodiabetes + usoinsulina)
```

```
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0892 -1.0246 -0.0327  0.9884  5.2002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.692291   0.545648  12.265 < 2e-16 ***
## sexo         -0.501599   0.120042  -4.179 3.35e-05 ***
## idade         0.020101   0.005839   3.443 0.000614 ***
## escolaridade -0.044630   0.014989  -2.978 0.003017 **
## imc           0.010711   0.014669   0.730 0.465559
## tempodiabetes 0.001615   0.000977   1.653 0.098765 .
## usoinsulina  -0.720681   0.117574  -6.130 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.483 on 633 degrees of freedom
## Multiple R-squared:  0.1125, Adjusted R-squared:  0.1041
## F-statistic: 13.37 on 6 and 633 DF, p-value: 2.661e-14
```

Método de Seleção de Modelos

backward, forward, stepwise

step(result, direction = "backward")

step(result, direction = "forward")

step(result, direction = "both")

Teste F para comparar a qualidade dos modelos com e sem a variável imc.

```
result2 <- update(result, ~. - imc)
```

```
summary(result2)
```

```
##
## Call:
## lm(formula = hemoglobina ~ sexo + idade + escolaridade + tempodiabetes +
##      usoinsulina)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0099 -1.0294 -0.0183  0.9833  5.2342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.9803192  0.3768565  18.522 < 2e-16 ***
## sexo         -0.5046332  0.1199261  -4.208 2.95e-05 ***
## idade         0.0200519  0.0058361   3.436 0.000629 ***
## escolaridade -0.0446243  0.0149835  -2.978 0.003010 **
## tempodiabetes 0.0016617  0.0009745   1.705 0.088657 .
## usoinsulina  -0.7206811  0.1175744  -6.130 1.55e-09 ***
```

```
## usoinsulina    -0.7197458  0.1175238  -6.124 1.60e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.482 on 634 degrees of freedom
## Multiple R-squared:  0.1117, Adjusted R-squared:  0.1047
## F-statistic: 15.95 on 5 and 634 DF,  p-value: 8.062e-15
```

```
anova(result2, result)
```

```
## Analysis of Variance Table
##
## Model 1: hemoglobina ~ sexo + idade + escolaridade + tempodiabetes + usoinsulina
## Model 2: hemoglobina ~ sexo + idade + escolaridade + imc + tempodiabetes +
##      usoinsulina
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      634 1392.8
## 2      633 1391.6  1    1.1721 0.5331 0.4656
```

- A comparação dos modelos indica que não existem indícios para rejeitar a hipótese nula de igualdade de qualidade dos modelos.
- Os modelos são semelhantes escolhendo-se, portanto, o modelo mais simples, pelo princípio da parcimônia.

Teste da razão de verossimilhança

```
#programa auxiliar
```

```
calc.estat.mod.comp.MRNLH <- function(result){ X <- model.matrix(result)
  logLikr <- logLik(result)
  n <- nrow(X)
  p <- ncol(X) #+ 1
  AICe <- AIC(result)
  BICe <- BIC(result)
  m2loglik <- -2*logLikr
  resultE <- c(AICe,BICe,m2loglik)
  names(resultE)<- c("AIC","BIC","-2log.lik")
  return(resultE)
  cat(estatisticac,resultE,"\n")
}
```

```
library(xtable)
```

```
m.estat <- cbind(rbind(AIC(result),BIC(result),
  as.numeric(logLik(result))),rbind(AIC(result2),
  BIC(result2),as.numeric(logLik(result2))))
TRV <- -2*(as.numeric(logLik(result)) - as.numeric(logLik(result2)))
pvalor <- 1-pchisq(TRV,df=1)

estmodcom1 <- calc.estat.mod.comp.MRNLH(result)
estmodcom2 <- calc.estat.mod.comp.MRNLH(result2)

m.estat <- cbind(estmodcom1,estmodcom2,pvalor)
xtable(m.estat)
```

```
## % latex table generated in R 4.4.1 by xtable 1.8-4 package
```

```

## % Thu Feb 13 16:38:10 2025
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrr}
## \hline
## & estmodcom1 & estmodcom2 & pvalor \\
## \hline
## AIC & 2329.35 & 2327.89 & 1.00 \\
## BIC & 2365.04 & 2359.12 & 1.00 \\
## -2log.lik & 2313.35 & 2313.89 & 1.00 \\
## \hline
## \end{tabular}
## \end{table}

```