

EXERCÍCIO 4

Na base de dados no arquivo “titanic_data.csv” deseja-se analisar quais são as variáveis que estão associadas com o a sobrevivência no naufrágio do Titanic. Pede-se:

a) Retire uma amostra de 300 observações utilizando a sintaxe abaixo:

```
# Amostra
set.seed(02051975) # COLOQUE O SEU DIA MÊS E ANO DE NASCIMENTO NA SET.SEED
base1 = base[sample(nrow(base), 300),]
```

- b) Classifique as variáveis independentes que forem qualitativas como fatores.
- c) Utilize o método de seleção de variáveis *Forward* para determinar o modelo final. A sintaxe é bem parecida com o método realizado na regressão linear múltipla.
- d) Determine as razões de chances com os intervalos de confiança para as variáveis selecionadas no modelo e faça a interpretação das razões de chances.

Variáveis	Categorias
Survival (var. dependente)	Sobreviveu ao naufrágio (1=Sim, 0=Não)
Pclass	Tipo de classe no Titanic (1=1ª classe, 2=2ª classe, 3=3ª classe)
Sex	Sexo do passageiro
Age	Idade em anos
Sibsp	Número de familiares a bordo do Titanic (irmãos, cônjuge)
Parch	Número de filhos a bordo do Titanic
Fare	Tarifa do passageiro (x\$1.000)
Embarked	Porto que embarcou (C=Cherbourg, Q=Queenstown, S=Southampton)

Observação: Em um arquivo PDF coloque a sintaxe do R, os resultados, as conclusões de cada tópico e poste no Moodle em EXERCICIO4, até às 13:00hs do dia 12/11 (terça-feira).