**FUNDAÇÃO GETÚLIO VARGAS**

**MBA Executivo em Business Analytcs e Big Data**

**Unidade Paulista**

**DISCIPLINA**

**Inferência Estatística**

**PROFESSOR  
Dr. João Rafael Dias Pinto**

**TRABALHO INDIVIDUAL**

**Análise Exploratória**

**ALUNA**Patrícia Yoshie Kishi Bueno

A57854913

São Paulo/SP

12/07/2020

**1. INTRODUÇÃO**

Trabalho individual realizado para avaliação dos conhecimentos adquiridos em inferência estatística utilizando o *RStudio* com *base R* e pacotes como *readxl, dplyr, summarytools e car* para a análise exploratória e descritiva dos dados e testes de hipóteses para a resolução do problema*.*

**2. O PROBLEMA**

**1** Análise do salário inicial de *MBA*

Vários são os cursos de *MBA* oferecidos, promovendo a excelência de seus programas, a qualidade do corpo docente e outros fatores como melhores salários.

Marie Daer, uma aspirante a *MBA* interessada nos salários iniciais dos graduados, montou através de uma pesquisa um *data frame* com os dados de uma escola três meses após a formatura dos alunos de 2012. Na pesquisa contém informações sobre a satisfação com o *MBA* e o salário inicial. As respostas foram adicionadas as informações deles já existentes no programa.

Com os dados, Marie, tenta esclarecer algumas perguntas que a ajudariam a decidir se matricula ou não no programa de *MBA* dessa escola:

- Qual o salário inicial? O sexo ou a idade influenciam no valor? A nota do *GMAT* faz diferença dado que a sua é baixa devido o inglês não ser sua língua nativa?

- Os alunos gostaram do curso?

**1** Resumo da fonte: MBA STARTING SALARIES Copyright © 2012, Richard Ivey School of Business Foundation

**3. ANÁLISE DOS DADOS**

A base de dados que contém as informações de salário inicial por ano, pontuação *GMAT* e outras informações sobre o desempenho de cada aluno é composta por 274 observações com 13 variáveis e valores *missing* presentes nas variáveis *salary* representados com os códigos 998 (não responderam a pesquisa) e 999 (não informaram); e na variável *satis* a ocorrência do código 999 também se faz presente.

Continuando a observação da variável *salary* com a identificação de valores zerados, dado o contexto das informações, foi entendido que estes podem ser tratados como formandos que não estavam trabalhando no momento da pesquisa. Então, estes juntos com os valores *missing* foram descartados da amostra.

O salário com valor 220000, considerado *outlier* também foi retirado da amostra para não dispersar as análises.

**4. OBJETIVO**

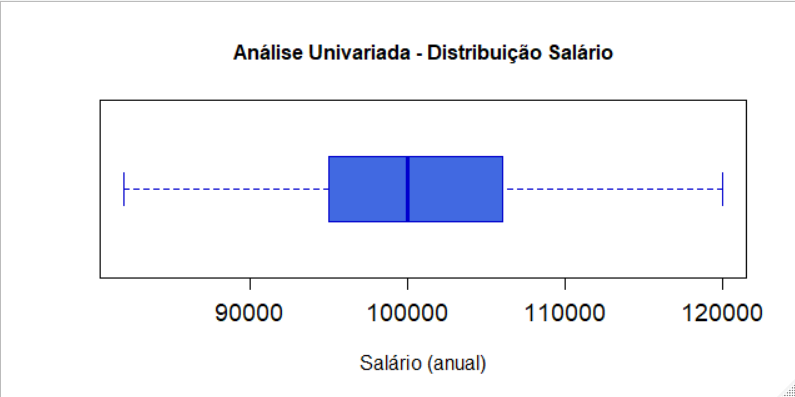
a) Após a formatura quanto Marie pode esperar de salário inicial.

b) Verificar se sexo, idade e notas do GMAT sofrem influência no valor do salário.

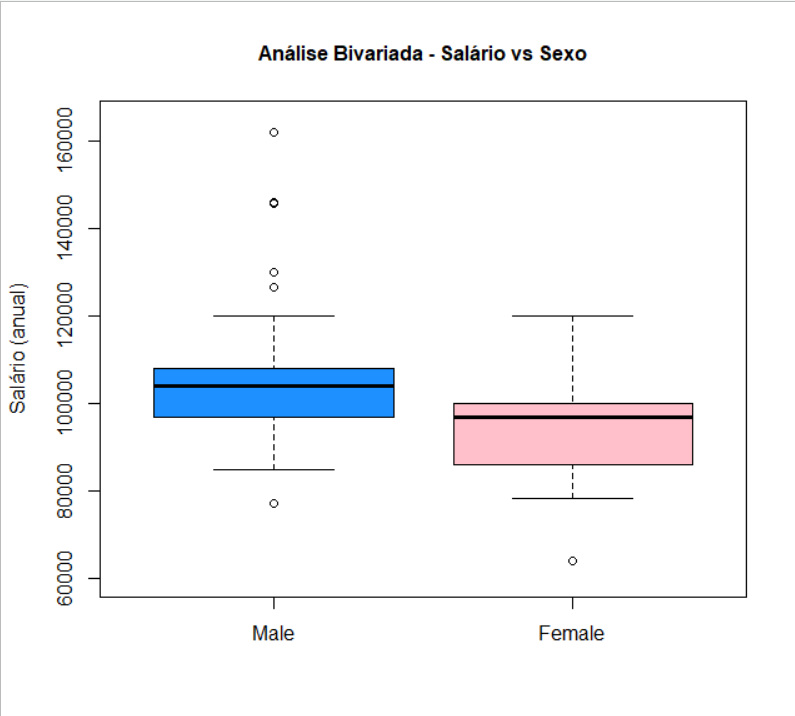
c) Recomendar ou não a matrícula para Marie.

**5. ESTUDO DO CASO**

Aproximadamente 50% dos salários iniciais estão entre 95.000 e 106.000 por ano.

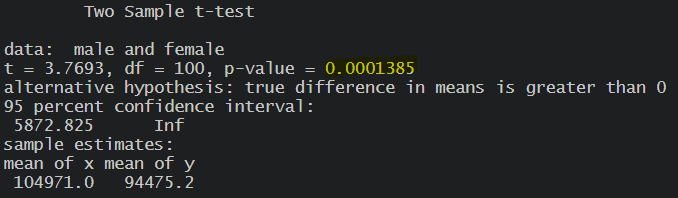


Com base nas medianas o sexo masculino tem salário maior do que o sexo feminino.

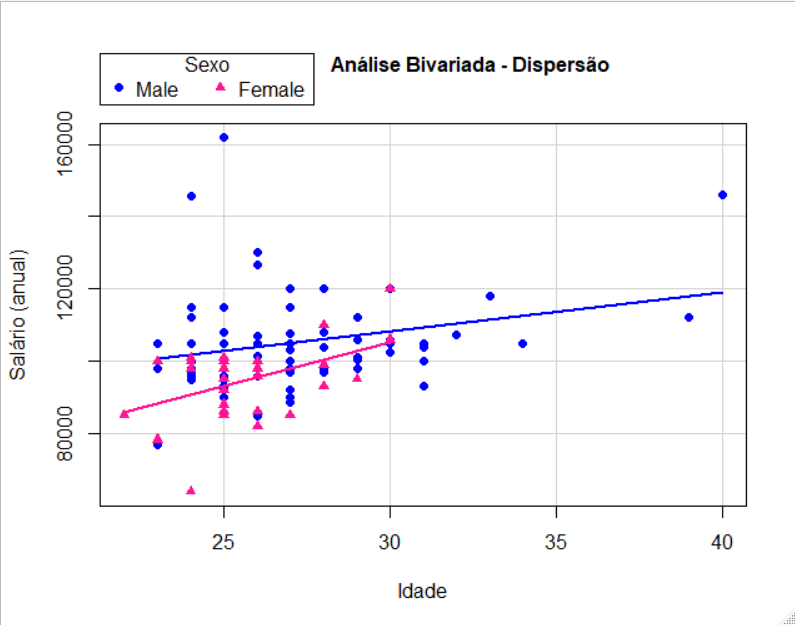


Observamos que a proporção da variância é similar, porém, no sexo feminino com valores mais baixo, constatada a hipótese alternativa ao aplicar o teste de hipótese *t* de *Student,* pelo resultado do p-valor 0.0001385 (< 0,05) que existe significância estatística de que a hipótese nula pode ser descartada, ou seja, existem evidências, na amostra analisada de que a média de salário do sexo masculino é maior do que o sexo feminino.

A seguir o resultado do cálculo do teste:

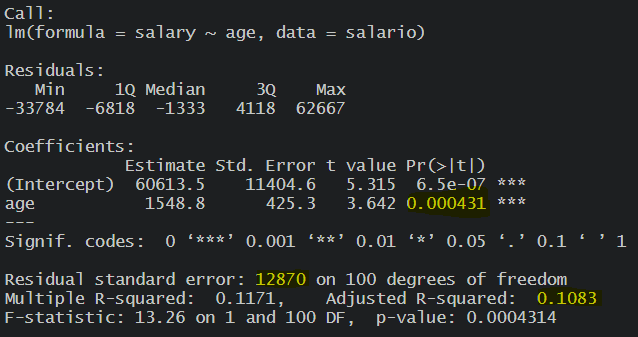
****

Comparando salário com idade observamos que existe uma relação linear direita com pouco grau de relacionamento na amostra, indicando uma relação pouco significante.

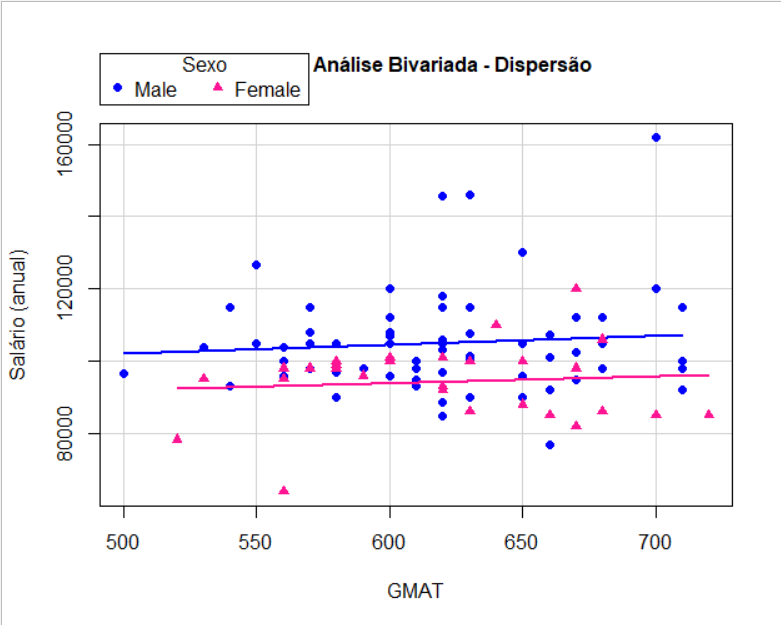


Porém, ao aplicarmos o teste de hipótese, conforme resultado abaixo, observamos que a relação encontrada é estatisticamente significante e com base na amostra, existe relação entre a idade e o salário, assim, conforme a idade aumenta, o salário também aumenta.

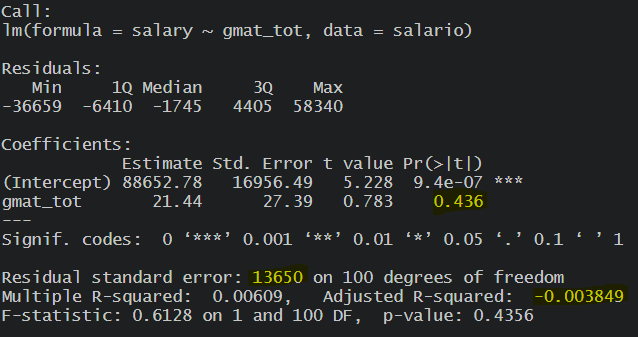
Como complemento, acrescento uma observação na análise: dado os valores apresentados pelo cálculo, as métricas p-valor e R2, além do valor um pouco alto de erro, para a afirmação exata sugere-se uma amostra maior.



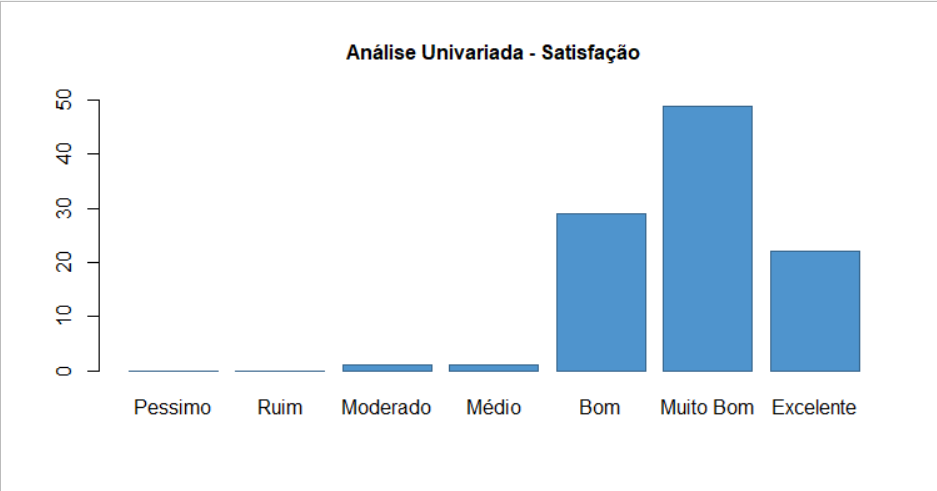
Já para a comparação de salário com a nota do GMAT está mais evidente que não existe nenhuma relação, a inclinação da reta quase horizontal.



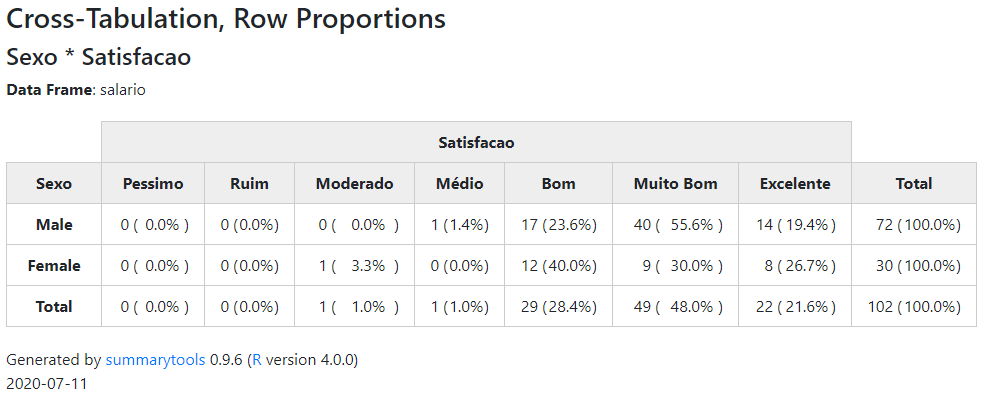
Assim, pelo resultado do teste de hipótese abaixo, confirmamos que a relação não é estatisticamente significante e com base na amostra, não existe relação entre a nota do GMAT com o salário.



Quanto ao nível satisfação dos formandos com relação ao programa de *MBA*, todos estão satisfeitos, conforme apresentação abaixo.

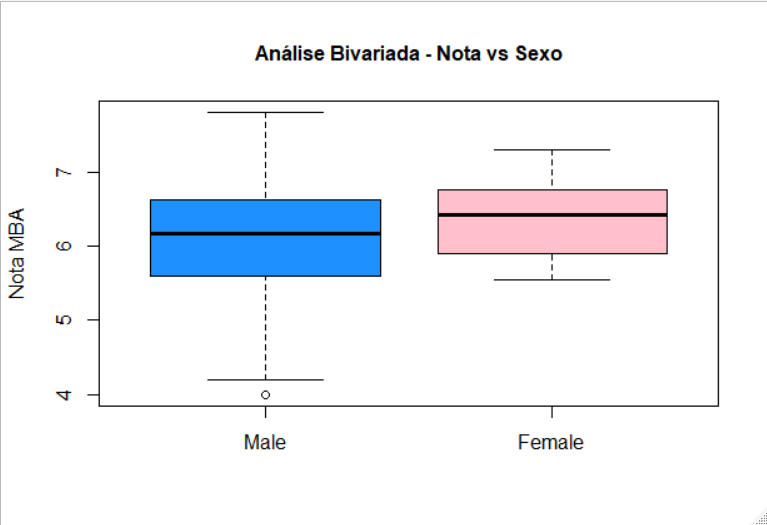


Pela representação na tabela cruzada, apenas 2% dos formandos não ficaram satisfeitos com o programa.

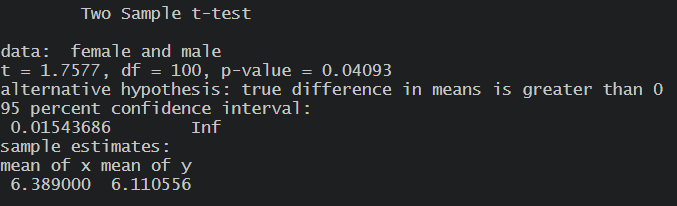


Assim, podemos recomendar a Marie a se inscrever no programa de *MBA,* dado o r*e*sultado apresen*t*ado pelas análises das amostras.

A amostra ainda indica que a média das notas do sexo feminino é maior do que a média do sexo masculino.



Provado pelo teste de hipótese que descarta a hipótese nula pelo p-valor 0.04093.

****

Tudo indica que Marie pode aproveitar muito bem o programa de MBA e ter grandes chances de um bom salário, sem receios da baixa nota no GMAT e não ter o inglês como sua língua nativa.