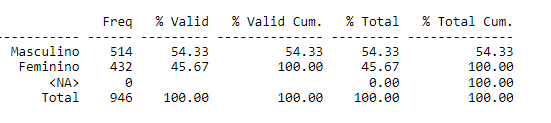
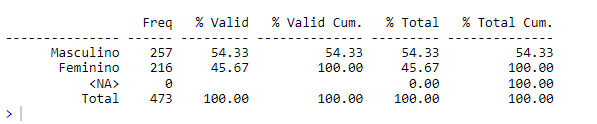
1. Tire uma tabela de frequência usando a função freq (summarytools) na variável Sexo. Quantos homens e quantas mulheres têm no arquivo?  
   
2. Remova os casos duplicados (olha para linha toda). E coloque esse arquivo limpo dentro de um outro arquivo chamado A. Para isso se utiliza a função unique no arquivo inteiro.

> A <- unique(Cadastral)

1. No arquivo A. Tire uma frequência usando a função freq (summarytools) na variável Sexo. Quantos homens e quantas mulheres possui o arquivo A? Precisa ser diferentes da primeira pergunta.



1. Crie uma variável data atual (utilize a função Sys.date) e acrescenta essa variável no arquivo A.

> A$DataAtual <- Sys.Date()

1. Verifique se a variável salario é numérica.

> is.numeric(A$salario)

[1] TRUE

1. Mostre o mínimo e o máximo da variável salario.  
     
   
2. Crie uma variável faixa de salario (função cut) com as seguintes quebras: 1574, 3000, 5000, 7000, 13500.  
     
   > faixa\_salario <- cut(A$salario,

+ breaks = c(1574,3000,5000,7000,13500),

+ labels = c("D", "C", "B", "A"))

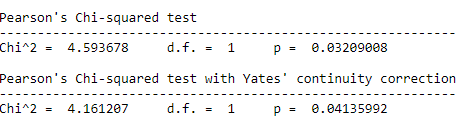
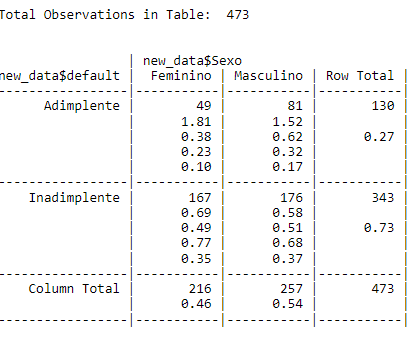
1. Abra o arquivo transacional e chama ele de B.

> B <- read\_excel("Transacional.xlsx")

1. Faça a união do arquivo A,B usando left join (função merge –packages base) do R. Consulte a função merge all.x  
     
   > new\_data = merge(A, B, all.x = TRUE)
2. Crie uma variável comprometimento de renda usando as variáveis ValorEmprestimo e Salario. Para isso utilize a expressão. (ValorEmprestimo / salario). Quantas variáveis ficaram no arquivo?   
     
   > new\_data$comprometimento = (new\_data$ValorEmprestimo / new\_data$salario)

> ncol(new\_data)

[1] 17

1. Faça um tabela cruzada usando duas variáveis categóricas (escolha do aluno) usando o packages gmodels. Explique a tabela.  
     
   Linha 1 (N) - Número de registros onde as duas variáveis comparadas se cruzam.

Linha 2 (contribuição do chi-quadrado) - Significa que quanto mais alto ele for, significa que é maior a diferença da frequência ocorrida do que era esperada pela tabela.

Linha 3 - O número de ocorrências em relação ao número de linhas. (Frequencia relativa)

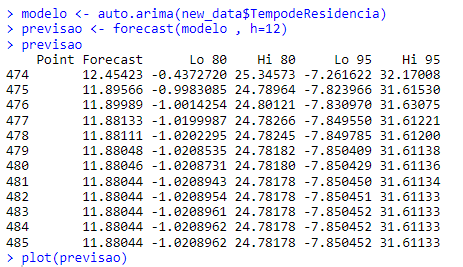
Linha 4 - O número de ocorrências em relação ao número de colunas.

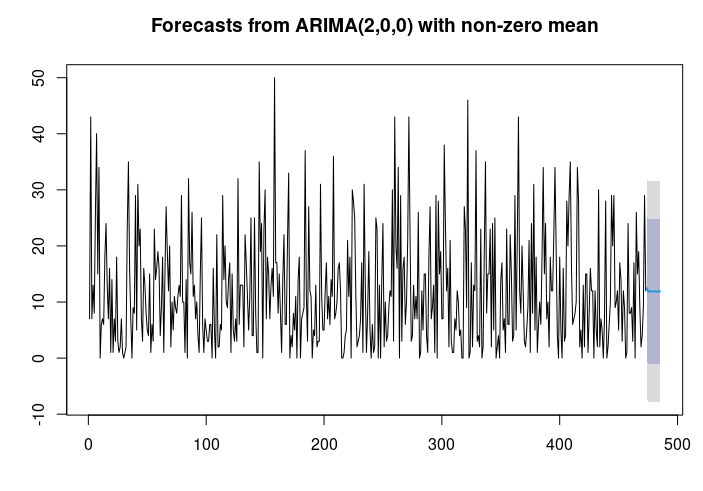
Linha 5 - o número de ocorrências em relação a tabela total.

Teste do chi quadrado: SE p-valor > 0.5 aceita H0.

SE p-valor < 0.5 Rejeita H0 e aceita H1.

Se a H1 for aceita, significa que as variáveis possuem correlação.

1. Escolha uma variável de seu preferencia e faça um previsão no tempo utilizando o packages forecast.  
   



1. Porque devemos criar novas variáveis?

Devemos criar novas variáveis para tomarmos nota do que foi concluído sobre o conjunto de dados atual e realizar estudos mais aprofundados do tema que está sendo estudado. Também é necessário gerar novas variáveis a partir das atuais para treinar algum modelo específico.

1. Faça um gráfico utilizando o packages ggplot2 e plotly. O aluno pode escolher qualquer variável.

> grafico\_barras <- ggplot(new\_data, aes(x=default, fill=Sexo))+

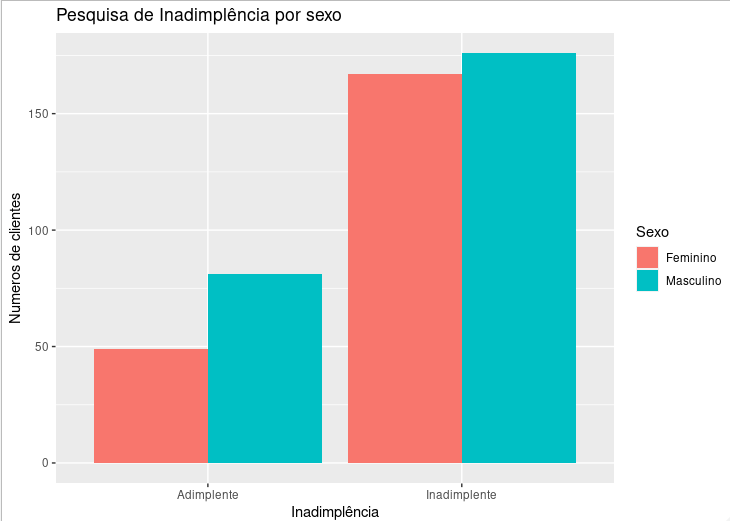
+ geom\_bar(position = 'dodge')+

+ xlab("Inadimplência")+

+ ylab("Numeros de clientes")+

+ ggtitle("Pesquisa de Inadimplência por sexo")

> grafico\_barras



**Plot com plotly**

fig <- plot\_ly(x = ~new\_data$salario, type = "histogram")

> fig

