

## PROBABILIDADE E INFERÊNCIA ESTATÍSTICA

### TRABALHO DO GRAU A 2021/2 – Peso 2,0

Este trabalho vale 2,0 pontos do Grau B e pode ser feito em duplas ou em trios. Deve ser elaborado um relatório que contenha os *Resultados* (com as interpretações) das **análises de estatística descritiva** realizadas para o banco de dados fornecido.

A data final de postagem do Trabalho do GA é no dia **30 de setembro de 2021**.

Entendendo o estudo...

A empresa *XYZ Corporation Internacional*, uma revendedora de automóveis de luxo começou as suas atividades em 2016 e atua nos 4 estados da região Sudeste (Espírito Santo, Minas Gerais, Rio de Janeiro e São Paulo).



O gerente comercial da empresa solicitou à TI a extração de um banco de dados para que sejam analisadas variáveis relacionadas às vendas a fim de auxiliar o gerente a verificar o comportamento das vendas e tomar decisões com base em análises estatísticas. O banco de dados é um arquivo em Excel e salvo em .CSV separado por vírgula (disponível no Moodle com o nome de *VendaCarros.csv*) com os dados coletados de Vendas e CRM da empresa, o qual possui as seguintes variáveis:

Variáveis (colunas)	Descrição	Níveis
Fabricante	Fabricante do veículo	7 fabricantes
Estado	Estado onde foi realizada a venda	RJ, SP, MG, ES
ValorVenda	Valor de venda do veículo	Quantitativa
ValorCusto	Custo do veículo para a empresa	Quantitativa
TotalDesconto	Total de desconto fornecido sobre o preço de venda	Quantitativa
CustoEntrega	Custo de entrega do veículo ao proprietário	Quantitativa
CustoMaoDeobra	Custo de mão de obra (secretária, mecânico, etc)	Quantitativa
Ano	Ano de venda do veículo	2016, 2017, 2018, 2019, 2020
Lucro	Lucro das vendas	Quantitativa

O trabalho consiste em utilizar o software *RStudio* e/ou *Excel* para realizar análises de estatísticas descritivas que julgares pertinente para auxiliar o gestor a estudar o comportamento das vendas, ou seja, **realizar análises descritivas para as variáveis do banco de dados e interpretar os resultados**:

- Resumos numéricos para as variáveis quantitativas (média, mediana, desvio padrão, coeficiente de variação): *Custo de Mão de Obra*, *Custo de Entrega*, *Total de Desconto*, *Valor do Custo*, *Valor de Venda* e *Lucro*;
- Tabela de frequência (para as variáveis qualitativas e para a variável *Ano*): *Fabricante*, *Estado* e *Ano*.
- Gráficos: histograma para as variáveis quantitativas (*Custo de Mão de Obra*, *Custo de Entrega*, *Total de Desconto*, *Valor do Custo*, *Valor de Venda* e *Lucro*) e gráfico de colunas ou de pizza para as variáveis qualitativas (*Fabricante* e *Estado*) e gráfico de colunas para a variável *Ano*.

### **Lembre-se de interpretar os resultados e montar uma conclusão para o trabalho.**

OBS: Utilizar as dicas do TUTORIAL DE INSTALAÇÃO DO R e DO RSTUDIO e demais comandos que foram descritos no Material de Apoio PARTE I.

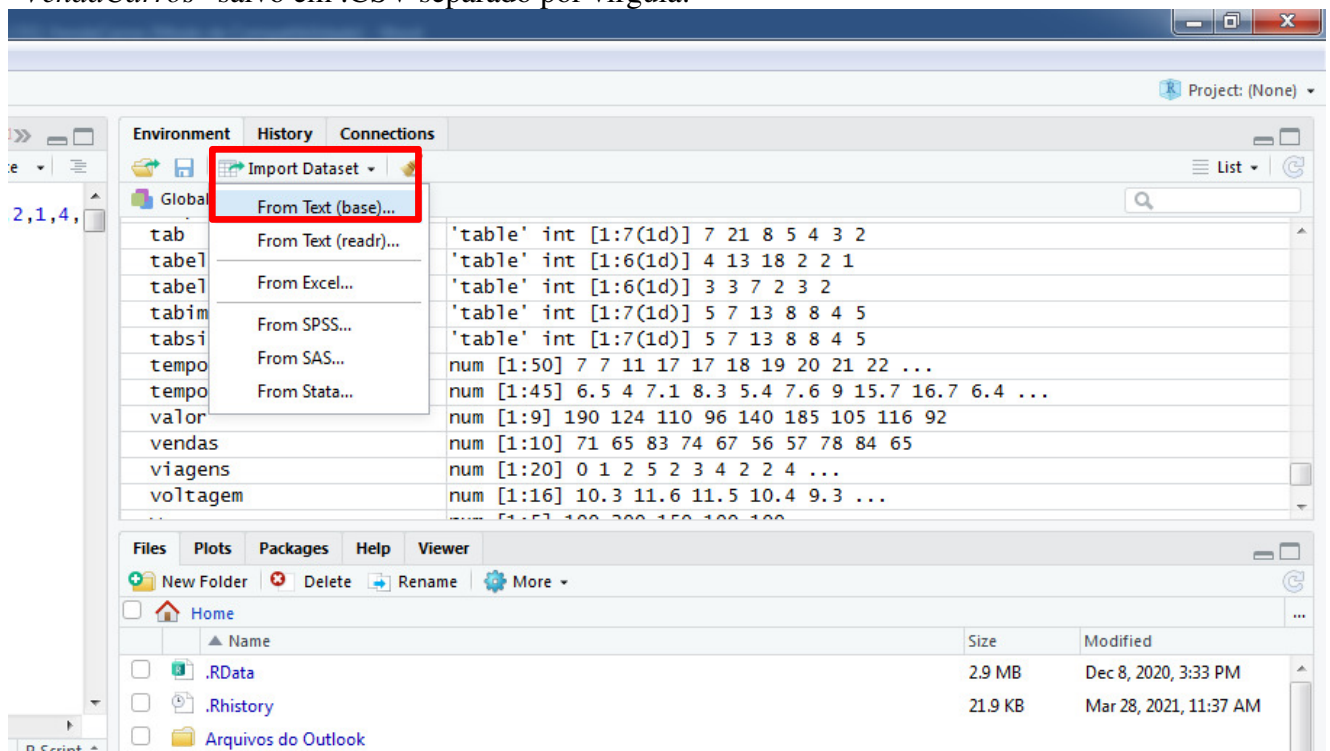
OBS: O site a seguir possui um repertório de comandos que podem ser realizados no R e podem ser úteis para a realização deste trabalho:

<http://dwoil.de/rexrepos/posts/sumProdRange.html>

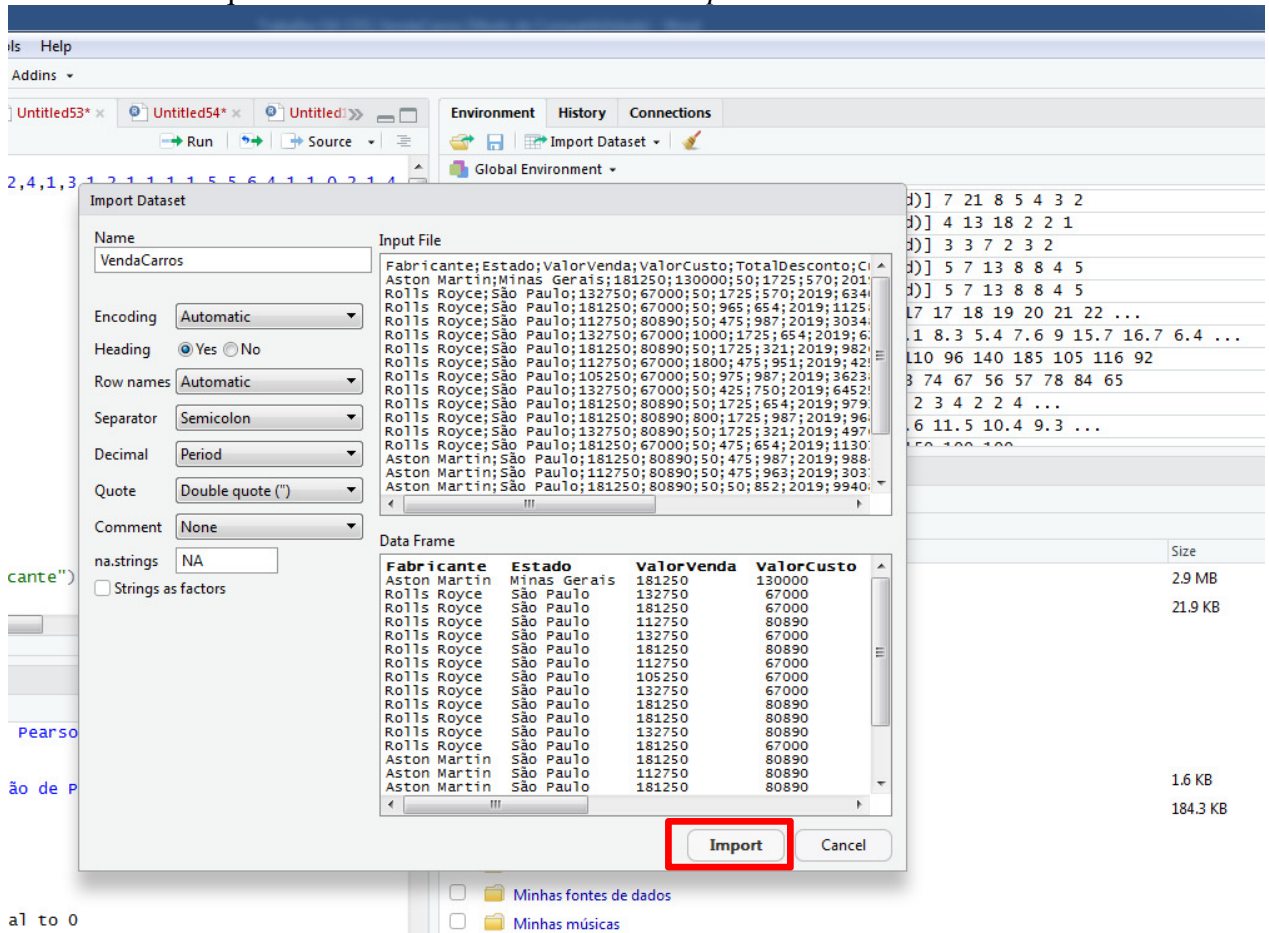
### **Sugestões de códigos que podem ser úteis para Trabalho do GA:**

Lembre-se de importar o banco de dados (salvo em .csv separado por vírgula) no *RStudio* seguindo o passo a passo para depois começar a trabalhar no Script (Tela 1):

- Ir em “*Import Dataset*”, selecionar *From Text (base)* e inserir o arquivo “*VendaCarros*” salvo em .CSV separado por vírgula:



- Seleccionar o arquivo “VendaCarros” e clicar em “Import”:



Import Dataset

Name: VendaCarros

Input File: Fabricante;Estado;ValorVenda;ValorCusto;TotalDesconto;CustoEntrega;CustoMaoDeObra;Ano;Lucro

Encoding: Automatic

Heading: Yes

Row names: Automatic

Separator: Semicolon

Decimal: Period

Quote: Double quote (")

Comment: None

na.strings: NA

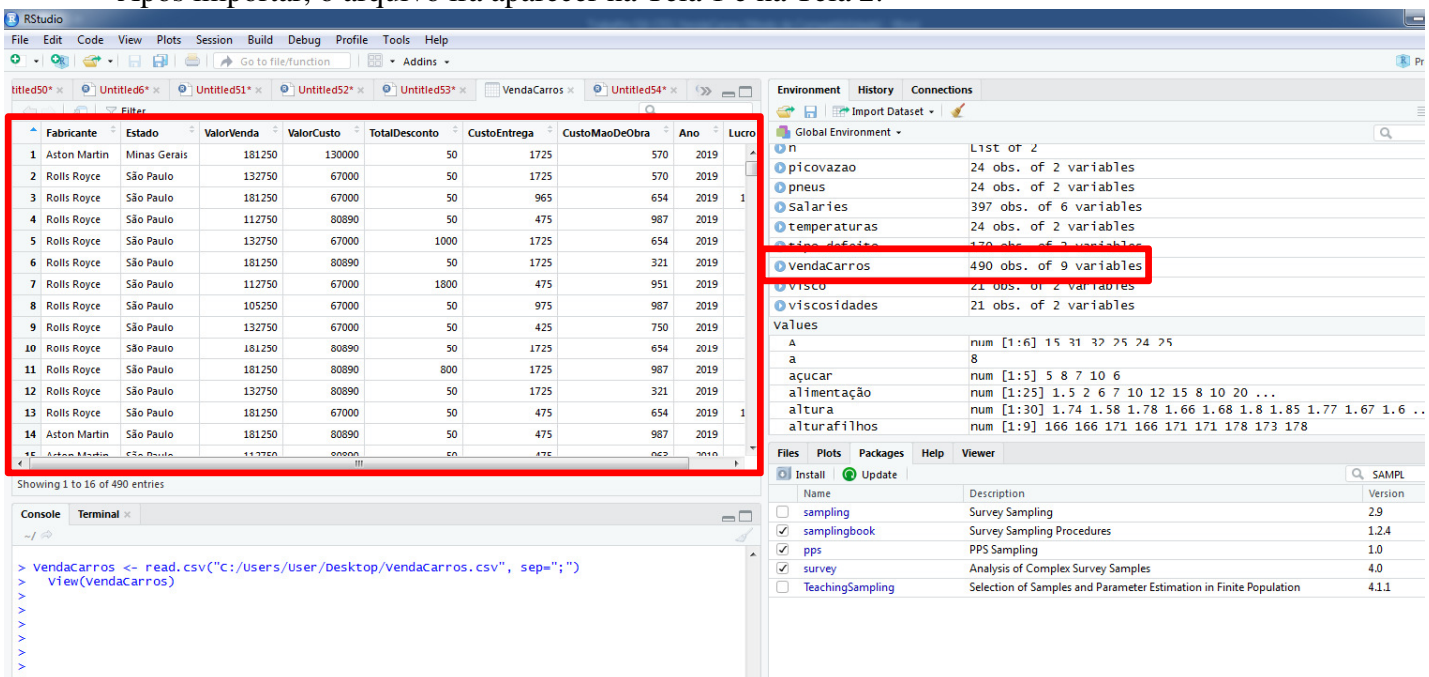
☐ Strings as factors

Data Frame

Fabricante	Estado	ValorVenda	ValorCusto	TotalDesconto	CustoEntrega	CustoMaoDeObra	Ano	Lucro
Aston Martin	Minas Gerais	181250	130000	50	1725	570	2019	
Rolls Royce	São Paulo	132750	67000	50	1725	570	2019	
Rolls Royce	São Paulo	181250	67000	50	965	654	2019	1
Rolls Royce	São Paulo	112750	80890	50	475	987	2019	
Rolls Royce	São Paulo	132750	67000	1000	1725	654	2019	
Rolls Royce	São Paulo	181250	80890	50	1725	321	2019	
Rolls Royce	São Paulo	112750	67000	50	475	951	2019	
Rolls Royce	São Paulo	105250	67000	50	975	987	2019	
Rolls Royce	São Paulo	132750	67000	50	425	750	2019	
Rolls Royce	São Paulo	181250	80890	50	1725	654	2019	
Rolls Royce	São Paulo	132750	67000	50	1725	987	2019	
Rolls Royce	São Paulo	181250	80890	50	475	654	2019	1
Rolls Royce	São Paulo	132750	67000	50	475	987	2019	
Aston Martin	São Paulo	181250	80890	50	475	987	2019	
Aston Martin	São Paulo	112750	80890	50	475	987	2019	
Aston Martin	São Paulo	181250	80890	50	475	987	2019	

Import Cancel

- Após importar, o arquivo irá aparecer na Tela 1 e na Tela 2:



Environment

Global Environment

VendaCarros 490 obs. of 9 variables

Values

A num [1:6] 15 31 37 25 24 25

a 8

açúcar num [1:5] 5 8 7 10 6

alimentação num [1:25] 1.5 2 6 7 10 12 15 8 10 20 ...

altura num [1:30] 1.74 1.58 1.78 1.66 1.68 1.8 1.85 1.77 1.67 1.6 ..

alturafilhos num [1:9] 166 166 171 166 171 171 178 173 178

Files Plots Packages Help Viewer

Install Update

NAME Description Version

sampling Survey Sampling 2.9

samplingbook Survey Sampling Procedures 1.2.4

pps PPS Sampling 1.0

survey Analysis of Complex Survey Samples 4.0

TeachingSampling Selection of Samples and Parameter Estimation in Finite Population 4.1.1

Console

```
VendaCarros <- read.csv("c:/Users/User/Desktop/VendaCarros.csv", sep=";")
View(VendaCarros)
```

```
#Pacote "descr"
install.packages("descr")
library(descr)

#Estado, Ano e Fabricante
freq(VendaCarros$Estado,ylim=c(0,300),main="Estado",ylab="Frequência")
freq(VendaCarros$Ano, ylim=c(0,300), main="Ano",ylab="Frequência")
freq(VendaCarros$Fabricante,ylim=c(0,200),main="Fabricante",ylab="Frequência")

piechart(VendaCarros$Fabricante, main="Fabricante")
piechart(VendaCarros$Estado, main="Estado")

tapply(VendaCarros$Ano, VendaCarros$Estado, freq)
tapply(VendaCarros$Estado, VendaCarros$Ano, freq)

#CUSTO MÃO DE OBRA:
summary(VendaCarros$CustoMaoDeObra)
sd(VendaCarros$CustoMaoDeObra)
CV1<-sd(VendaCarros$CustoMaoDeObra)/mean(VendaCarros$CustoMaoDeObra)
CV1

hist(VendaCarros$CustoMaoDeObra,nclass=7,ylab="Frequência",xlab="Custo
de Mão de Obra",main="Custo de Mão de Obra")

tapply(VendaCarros$CustoMaoDeObra,VendaCarros$Estado,summary)

#CUSTO ENTREGA:
summary(VendaCarros$CustoEntrega)
sd(VendaCarros$CustoEntrega)
CV2<-sd(VendaCarros$CustoEntrega)/mean(VendaCarros$CustoEntrega)
CV2

hist(VendaCarros$CustoEntrega,nclass = 4,ylab="Frequência",xlab="Custo
de Entrega",main="Custo de Entrega")

tapply(VendaCarros$CustoEntrega,VendaCarros$Estado,summary)

#TOTAL DESCONTO:
summary(VendaCarros$TotalDesconto)
sd(VendaCarros$TotalDesconto)
CV3<-sd(VendaCarros$TotalDesconto)/mean(VendaCarros$TotalDesconto)
CV3

hist(VendaCarros$TotalDesconto,nclass=5,ylim=c(0,500),ylab="Frequência
",xlab="Total do Desconto",main="Total do Desconto")

tapply(VendaCarros$TotalDesconto,VendaCarros$Estado,summary)

#VALOR CUSTO:
summary(VendaCarros$ValorCusto)
sd(VendaCarros$ValorCusto)
CV4<-sd(VendaCarros$ValorCusto)/mean(VendaCarros$ValorCusto)
CV4

hist(VendaCarros$ValorCusto,nclass=4,ylim=c(0,400),ylab="Frequência",
xlab="Valor do Custo",main="Valor do Custo")

tapply(VendaCarros$ValorCusto,VendaCarros$Estado,summary)

#VALOR VENDA:
summary(VendaCarros$ValorVenda)
sd(VendaCarros$ValorVenda)
CV5<-sd(VendaCarros$ValorVenda)/mean(VendaCarros$ValorVenda)
CV5

hist(VendaCarros$ValorVenda,nclass=4,ylim=c(0,300),ylab="Frequência",
xlab="Valor de Venda",main="Valor de Venda")
```

```
tapply(VendaCarros$ValorVenda,VendaCarros$Estado,summary)
```

```
#LUCRO:
```

```
summary(VendaCarros$Lucro)
```

```
sd(VendaCarros$Lucro)
```

```
CV6<-sd(VendaCarros$Lucro)/mean(VendaCarros$Lucro)
```

```
CV6
```

```
hist(VendaCarros$Lucro,nclass=4,ylim=c(0,350),ylab="Frequência",  
xlab="Lucro",main="Lucro")
```

```
tapply(VendaCarros$Lucro,VendaCarros$Estado,summary)
```