PROVA MÉTODOS ESTATÍSTICOS

Discentes: Julie Rhanna e Roberta Graser

**Questão 4: Um assistente social deseja saber o tamanho da amostra (n) necessário para determinar a proporção da população atendida por uma Unidade de Saúde, que pertence ao município de Seropédica. Não foi feito um levantamento prévio da proporção amostral e, portanto, seu valor é desconhecido. Ela quer ter 95% de confiança que sua o erro máximo da estimativa seja de 6%. Quantas pessoas necessitam ser entrevistadas?**

Resposta: 267 pessoas.

> #####Calculate sample size from proportion and confidence interval#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> SampleProportionCI(0.5, 0.12, 95)

Assumptions

P 0.5

Confidence interval 0.12

Confidence level 0.95

Estimated

Required sample size 267

**Questão 5: O banco de dados chamado “tbreal\_prova2020.xlsx” trata-se de uma avaliação retrospectiva de uma investigação de surtos de tuberculose em animais de grande porte (bovinos leiteiros e de corte, cervídeos e bisões) no Canadá entre os anos de 1985 e 1994. Para investigar fatores de risco para a propagação da tuberculose nos rebanhos. Registros detalhados da investigação epidemiológica do surto durante o período foram revisados e os dados processados em uma planilha. Em cada animal, a data mais provável em que a infecção entrou no rebanho foi determinada a partir dos resultados dos testes diagnósticos. O objetivo deste trabalho é investigar os efeitos da idade, sexo, peso e tipo de animal no acometimento da doença. Nota 1: Para atender a questões de confidencialidade e regulamentação, esses dados foram deliberadamente falsificados. Nota 2: Nesse arquivo, os dados se encontram na primeira planilha, na segunda planilha está o dicionário de variáveis.**

Roteiro de análise:

**- Verifique o tipo de cada variável no banco de dados**

IDADE: Qualitativa Nominal

SEXO: Qualitativa Nominal

PESO: Quantitativa Contínua

TIPO DO ANIMAL: Qualitativa Nominal

Nº DO ANIMAL: Qualitativa Ordinal

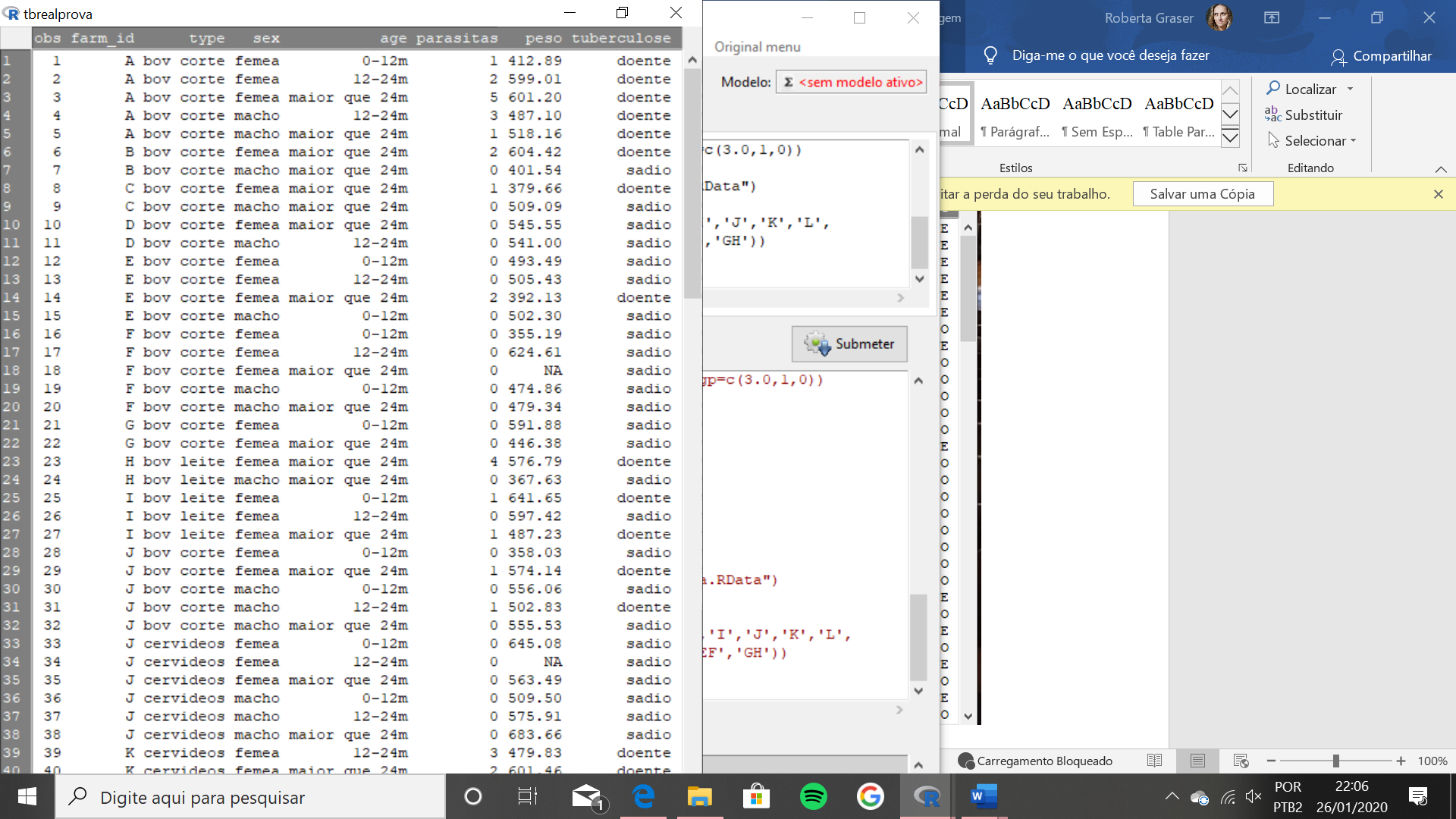
IDENTIFICAÇÃO DA FAZENDA: Qualitativa Nominal

PARASITA: Quantitativa Discreta

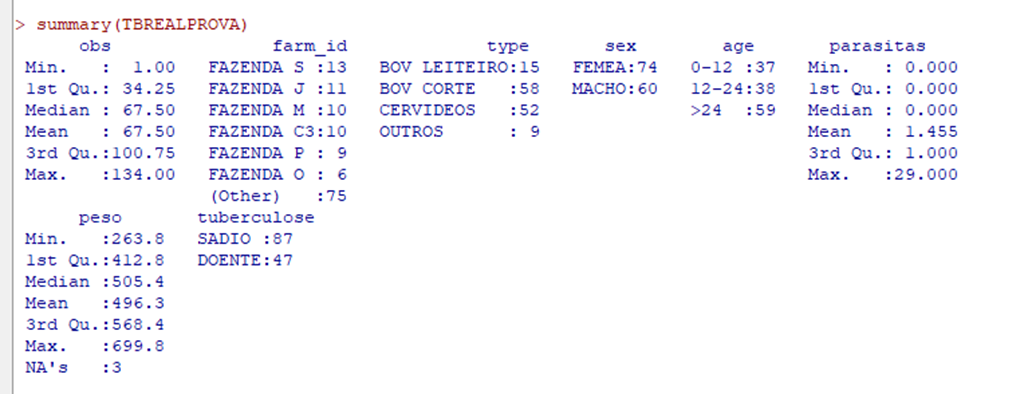
TUBERCULOSE: Qualitativa Nominal

**- Análise exploratória ou descritiva:**

**I) Transformar para fator e colocar os devidos *labels* das variáveis categóricas;**



**II) Fazer o sumario estatístico das variáveis quantitativas e as distribuições de frequências para as variáveis qualitativas;**

****

> numSummary(tbrealprova[,c("parasitas", "peso"), drop=FALSE], statistics=c("mean", "sd",

+ "se(mean)", "IQR", "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

**Mean sd se(mean) IQR cv 0% 25% 50% 75%**

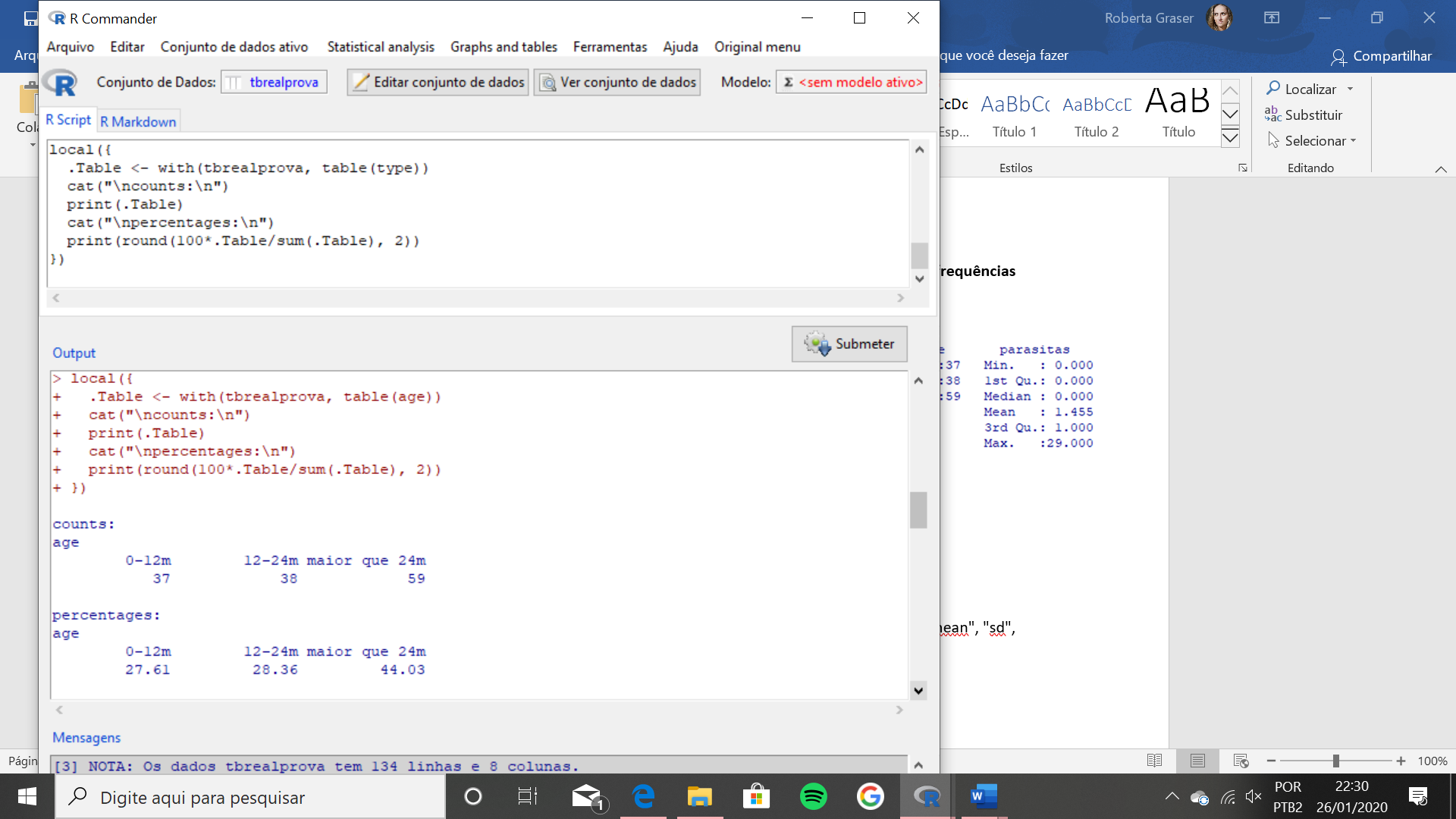
**parasitas** 1.455224 4.012416 0.346620 1.000 2.7572501 0.00 0.000 0.00 1.00

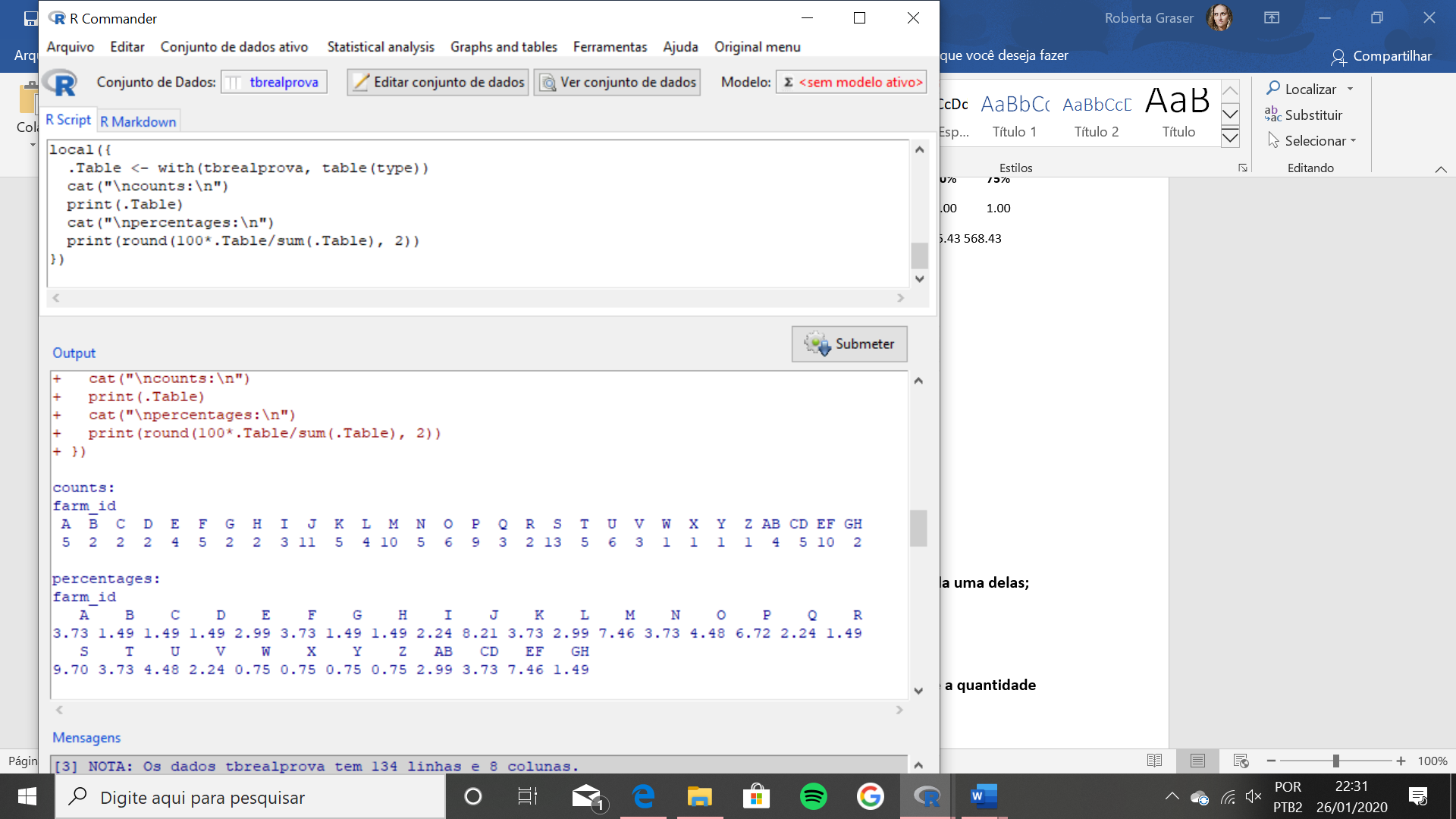
**peso** 496.296183 98.363228 8.594035 155.675 0.1981946 263.75 412.755 505.43 568.43

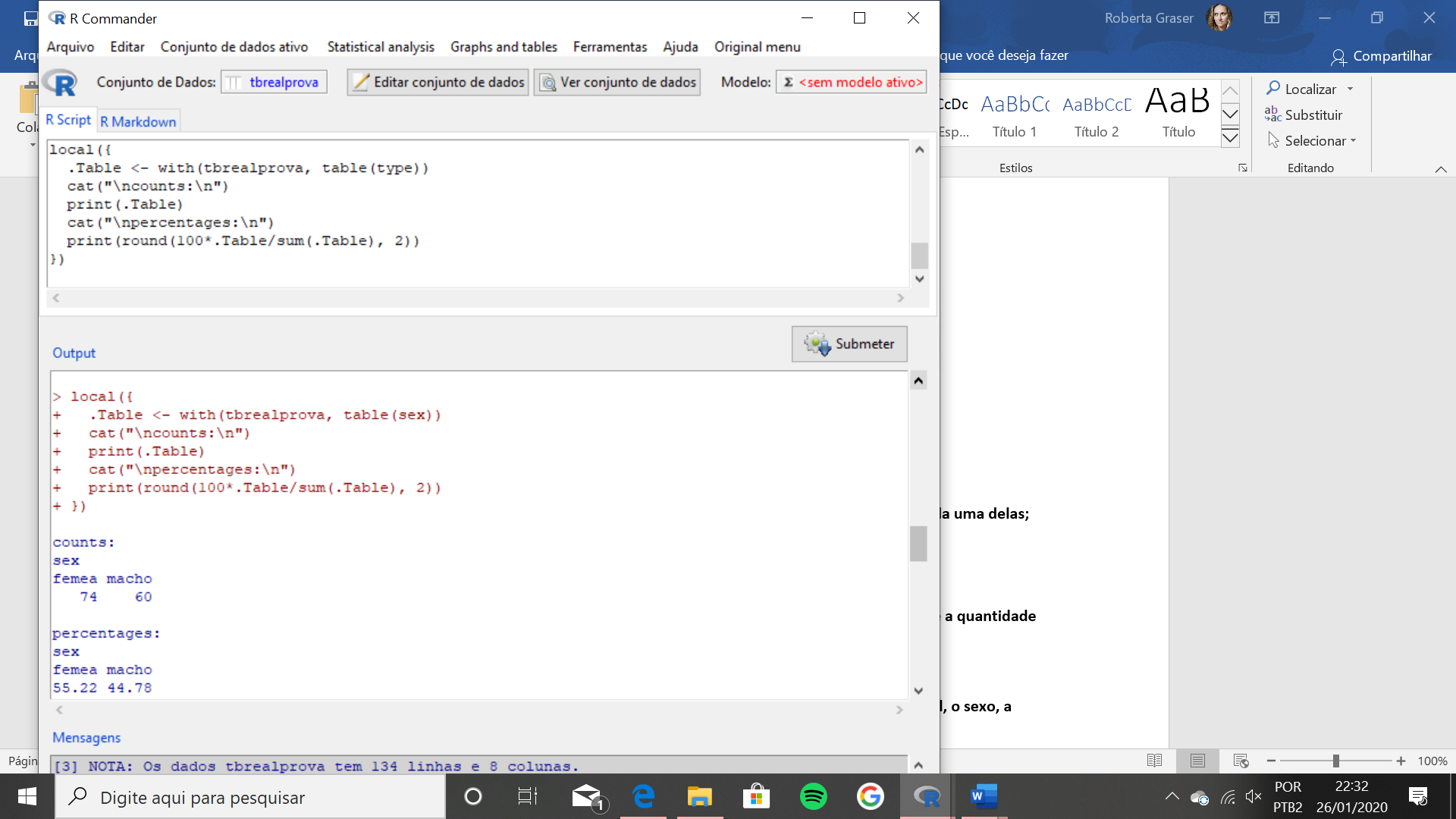
**100% n NA**

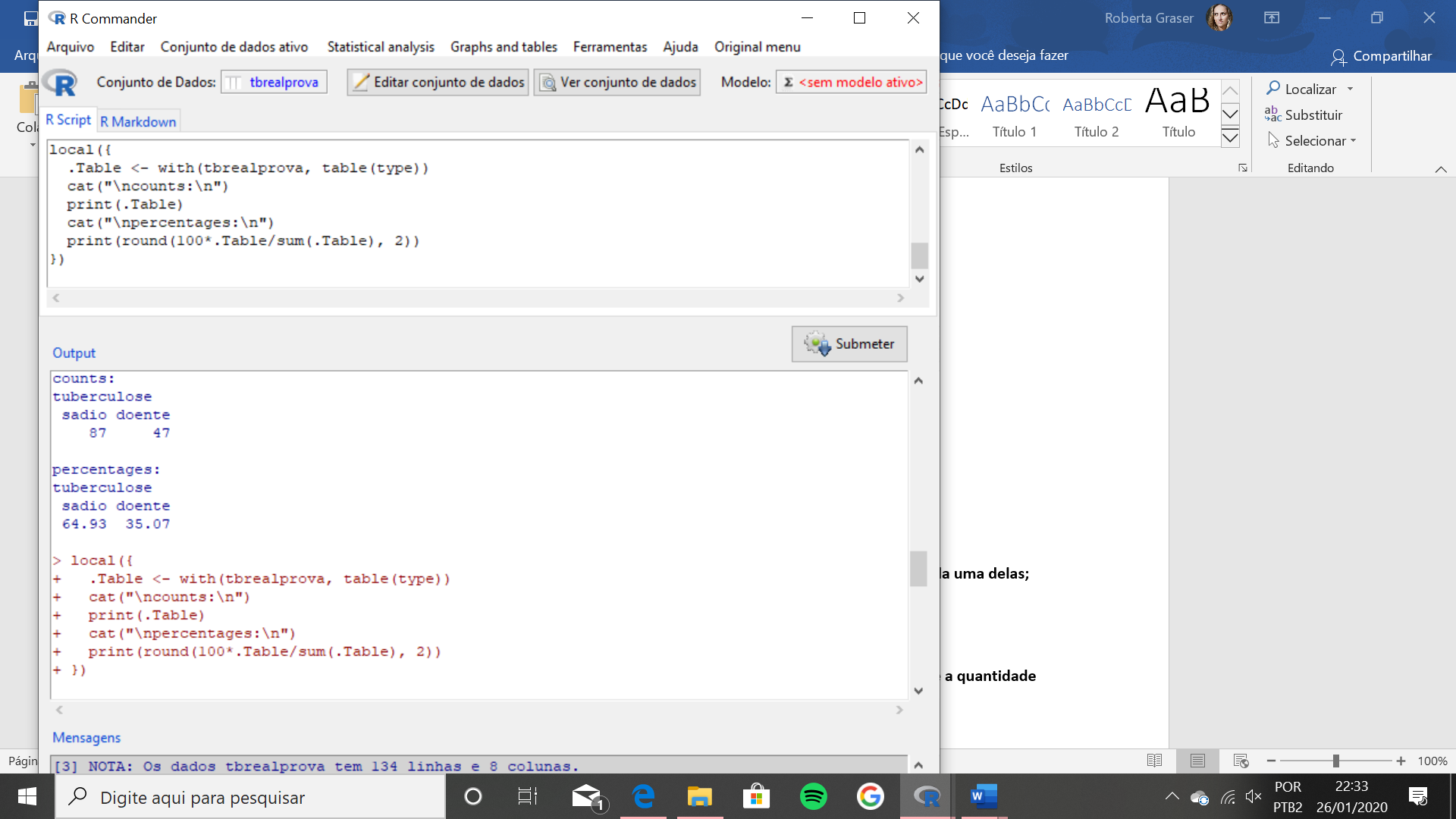
**parasitas** 29.00 134 0

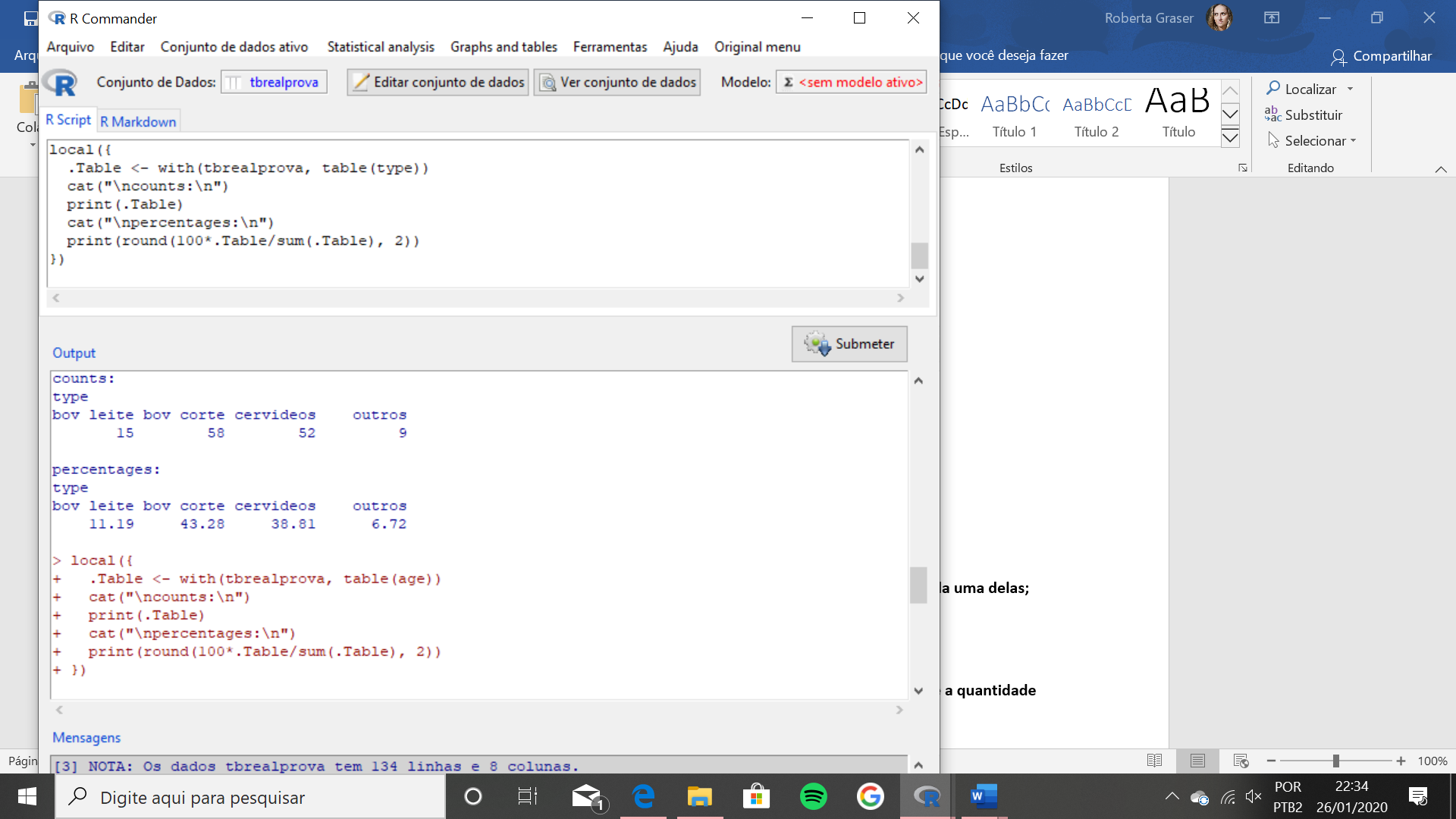
**peso** 699.81 131 3









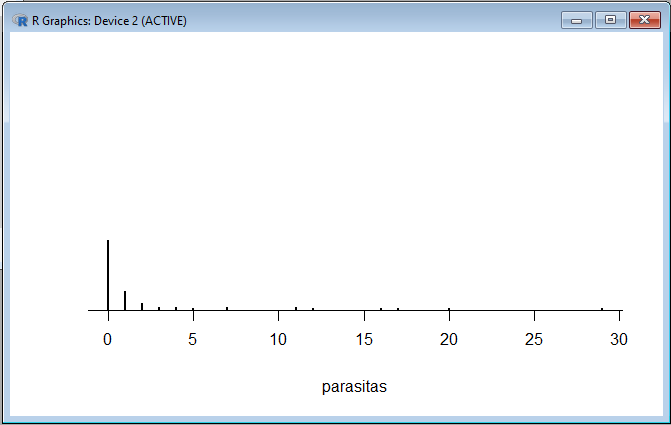


**III) Plotar os devidos gráficos das variáveis para mostrar as distribuições de cada uma delas;**

Optou-se pela utilização de gráficos de pontos e histograma, para variáveis quantitativas e gráficos em barras e em setor para representação das variáveis qualitativas .

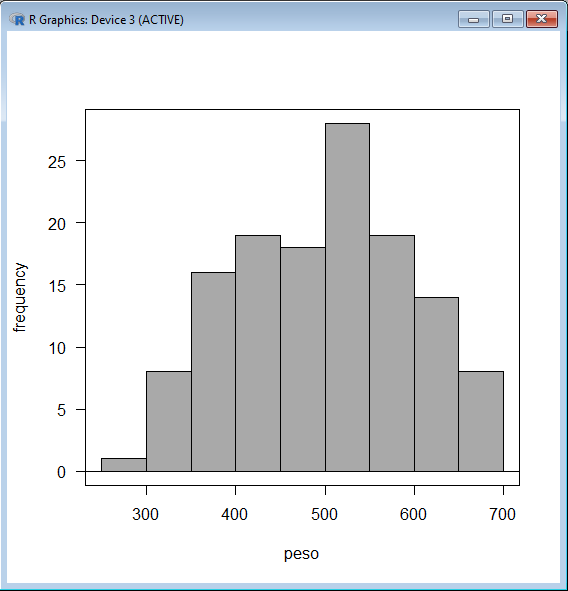
PARASITAS: Gráfico de pontos.

> with(prova, Dotplot(parasitas, bin=FALSE))



PESO: Histograma

> with(prova, Hist(peso, scale="frequency", breaks="Sturges", col="darkgray"))



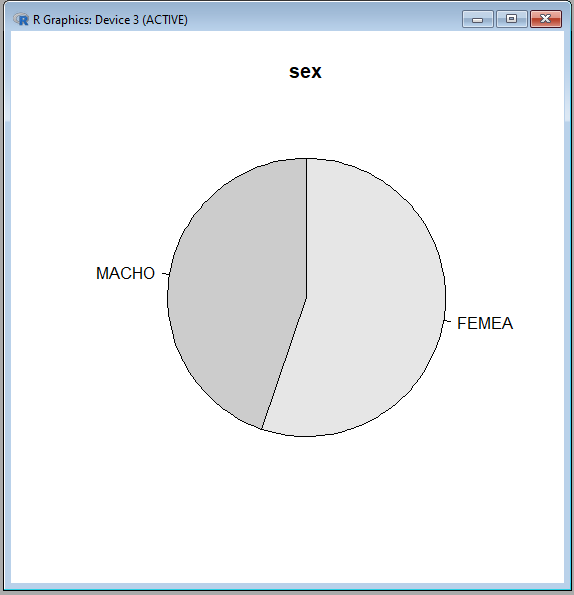
SEXO: Gráfico em setor

> #####Pie chart(Frequencies)#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> pie(table(prova$sex), labels=levels(factor(prova$sex)), main="sex", col=(gray(c(0.9,

+ 0.8))), clockwise=TRUE)

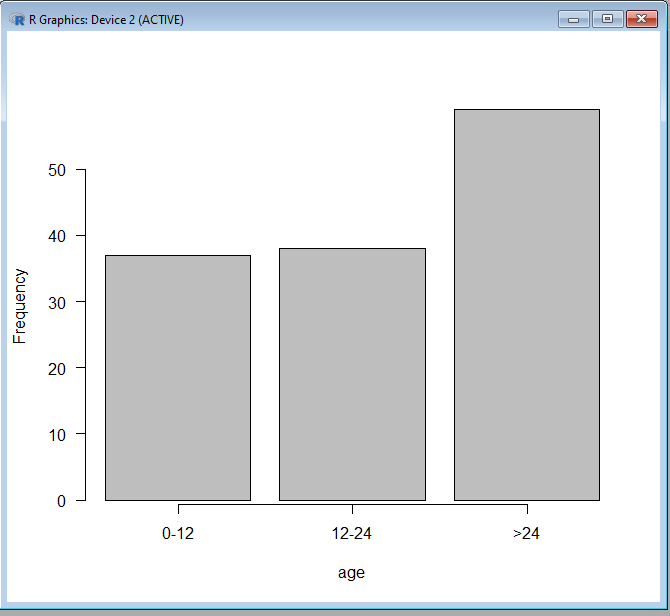


IDADE: Gráfico em Barras

#####Bar graph(Frequencies)#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> barplot(table(prova$age), xlab="age", ylab="Frequency", axis.lty=1)



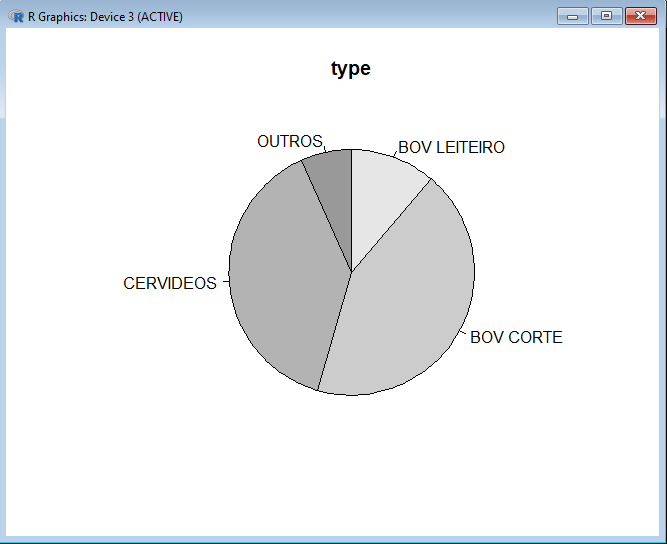
TIPO DE ANIMAL: Gráfico em Setor

> #####Pie chart(Frequencies)#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> pie(table(prova$type), labels=levels(factor(prova$type)), main="type",

+ col=(gray(c(0.9, 0.8, 0.7, 0.6))), clockwise=TRUE)



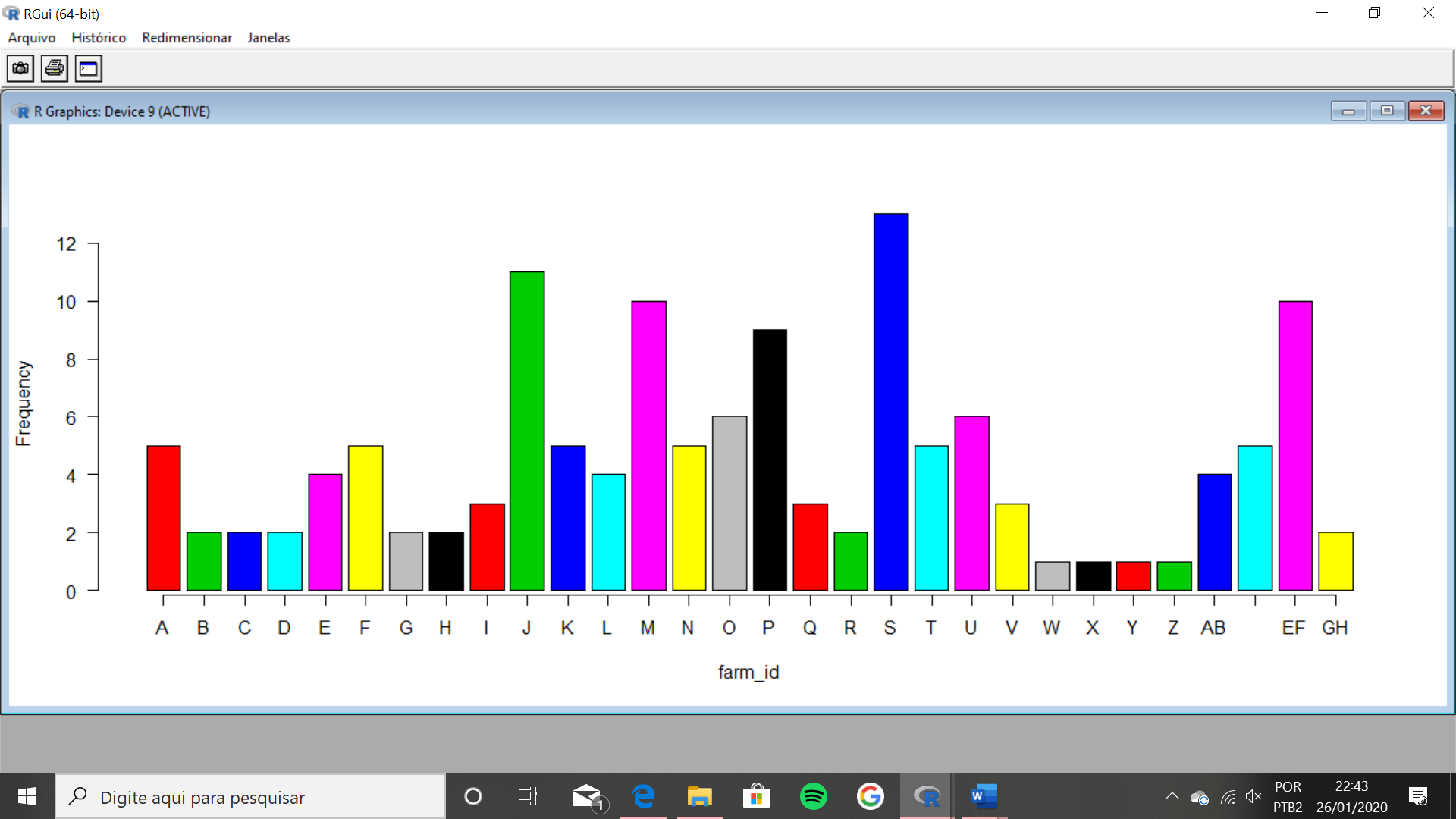
ID FAZENDA: Gráfico de Barras

> #####Bar graph(Frequencies)#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> barplot(table(prova$farm\_id), xlab="farm\_id", ylab="Frequency", col=c(2:31),

+ axis.lty=1)



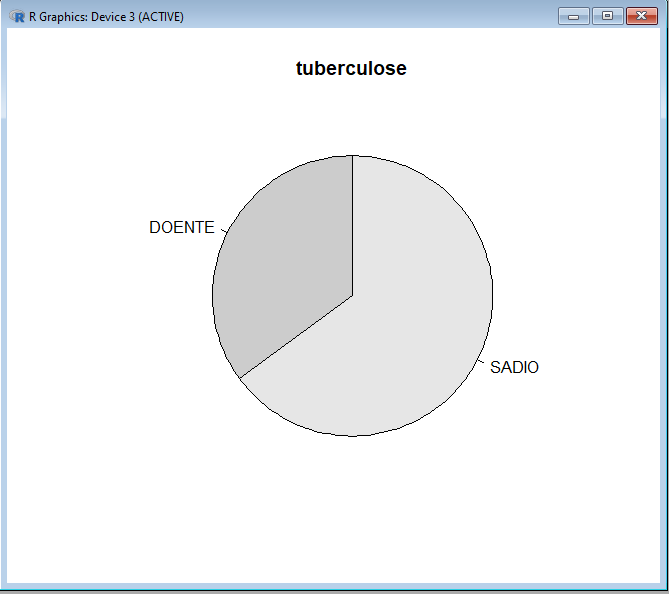
TUBERCULOSE: Gráfico em Setor

> #####Pie chart(Frequencies)#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> pie(table(prova$tuberculose), labels=levels(factor(prova$tuberculose)),

+ main="tuberculose", col=(gray(c(0.9, 0.8))), clockwise=TRUE)



**- Inferência Estatística:**

**Obs 2: Não se esqueça de testar a normalidade dos dados, para aplicar o teste estatístico correto. E também teste se existe heterogeneidade entre as variâncias.**

TESTES DE NORMALIDADE (para variáveis quantitativas)

> normalityTest(~peso, test="shapiro.test", data=tbrealprova)

Shapiro-Wilk normality test

data: **peso**

W = 0.98562, **p-value = 0.1847**

**O p-valor da variável Peso encontra-se dentro da normalidade.**

> normalityTest(~parasitas, test="shapiro.test", data=tbrealprova)

Shapiro-Wilk normality test

data: **parasitas**

W = 0.40502, **p-value < 2.2-16**

**O p-valor da variável Parasitas encontra-se fora da normalidade.**

Bartlett test of homogeneity of variances //

data: peso by tuberculose

> bartlett.test(peso ~ tuberculose, data=prova)

Bartlett test of homogeneity of variances

data: peso by tuberculose

Bartlett's K-squared = 0.3449, df = 1, **p-value = 0.557**

**Bartlett test p.value = 0.557**

**O p-valor de peso X tuberculose é maior que o nível de significância (0,05), não rejeita H0, sendo assim, conclui-se que as variâncias são constantes ou homogêneas.**

> (res <- bartlett.test(parasitas ~ tuberculose, data=tbrealprova))

Bartlett test of homogeneity of variances

data: parasitas by tuberculose

Bartlett's K-squared = Inf, df = 1, **p-value < 2.2e-16**

**Bartlett test p.value = 0**

**O p-valor neste caso é menor que o nível de significância (0,05), rejeita-se a hipótese nula (H0) em favorecimento da hipótese alternativa (H1) - neste caso assume-se a desigualdade das variâncias; o p-valor do teste de Bartlett foi menor que 0,05, assim há heterogeneidade das variâncias.**

1. **Verificar se a patologia está associada ao: tipo do animal, sexo, idade, peso e a quantidade de parasitas:**

**TUBERCULOSE X TIPO DO ANIMAL**

> #####Create two-way table and compare two proportions (Fisher's exact test)#####

> Fisher.summary.table <- NULL

> .Table <- NULL

> .Table <- xtabs(~type+tuberculose, data=tbrealprova)

> .Table

**Tuberculose (Percentages)**

**type sadio doente**

**bov leite** 10 (11.5%) 5 (10.6%)

**bov corte** 38 (43.7%) 20 (42.6%)

**cervídeos** 33 (37.9%) 19(40.4%)

**outros**  6 (6.9%) 3 (6.4%)

**Total** 100.0% 100.0%

**Count** 87.0 47.0

Pearson's Chi-squared test

X-squared = 0.089808, df = 3, **p-value = 0.993**

> Fisher.summary.table

tuberculose=sadio tuberculose=doente **Chisq.p.value**

type=bov leite 10 5 **0.993**

type=bov corte 38 20

type=cervideos 33 19

type=outros 6 3

O p-value = 0.993 é maior que 0.05: nao rejeita-se H0, verificamos que as variáveis tipo de animal e a ocorrência de tuberculose não apresentam associação.

**TUBERCULOSE X SEXO**

> #####Create two-way table and compare two proportions (Fisher's exact test)#####

> Fisher.summary.table <- NULL

> .Table <- NULL

> .Table <- xtabs(~sex+tuberculose, data=tbrealprova)

> .Table

**Tuberculose (Percentages)**

**sex sadio doente**

**femea** 43 (49.4%) 31 (66%)

**macho** 44 (50.6%) 16 (34%)

**Total** 100.0% 100%

**Count** 87.0 47

Pearson's Chi-squared test

X-squared = 3.3729, df = 1, **p-value = 0.06628**

> Fisher.summary.table

tuberculose=sadio tuberculose=doente **Chisq.p.value**

sex=femea 43 31 **0.0663**

sex=macho 44 16

O p-value = 0.0663 é maior que 0.05: não rejeita-se H0, verificamos que as variáveis sexo e a doença tuberculose não apresentam associação.

**TUBERCULOSE X IDADE**

> #####Create two-way table and compare two proportions (Fisher's exact test)#####

> Fisher.summary.table <- NULL

> .Table <- NULL

> .Table <- xtabs(~age+tuberculose, data=tbrealprova)

> .Table

**Tuberculose (Percentages)**

**age sadio doente**

**0-12m** 35 (40.2%) 2 (4.3%)

**12-24m** 23 (26.4%) 15 (31.9%)

**maior que 24m** 29 (33.3%) 30 (63.8%)

**Total** 99.9% 100.0%

**Count** 87.0 47.0

Pearson's Chi-squared test

X-squared = 21.071, df = 2, **p-value = 0.00002658**

> Fisher.summary.table

tuberculose=sadio tuberculose=doente **Chisq.p.value**

age=0-12m 35 2 **0.0000266**

age=12-24m 23 15

age=maior que 24m 29 30

O p-valor = 0.0000266 é menor que 0.05: rejeita-se H0, as variáveis apresentam associação, ou seja, o idade pode estar relacionada com a tuberculose, animais com mais de 24 meses tem mais tuberculose.

**TUBERCULOSE X PESO**

Pearson's Chi-squared test

data: .Table

X-squared = 131, df = 130, **p-value = 0.4589**

(A tabela é muito grande para colocar aqui!)

O p-value = 0.4589 é maior que 0.05: aceita-se H0, verificamos que as variáveis peso e a doença tuberculose não apresentam associação; Desta forma, assume-se que não existe correlação entre peso e ocorrência da tuberculose.

**TUBERCULOSE X QTD PARASITAS**

> #####Create two-way table and compare two proportions (Fisher's exact test)#####

> Fisher.summary.table <- NULL

> .Table <- NULL

> .Table <- xtabs(~parasitas+tuberculose, data=tbrealprova)

> .Table

**Tuberculose (Percentages)**

**parasitas sadio doente**

0 87 (100%) 0 (0%)

1 0 (0%) 23 (48.9%)

2 0 (0%) 8 (17.0%)

3 0 3 (6.4%)

4 0 3 (6.4%)

5 0 1 (2.1%)

7 0 2 (4.3%)

11 0 2 (4.3%)

12 0 1 (2.1%)

16 0 1 (2.1%)

17 0 1 (2.1%)

20 0 1 (2.1%)

29 0 1 (2.1%)

**Total**  100% 99.9%

**Count** 87 47.0

Pearson's Chi-squared test

X-squared = 134, df = 12, **p-value < 2.2-16**

> Fisher.summary.table

tuberculose=sadio tuberculose=doente **Chisq.p.value**

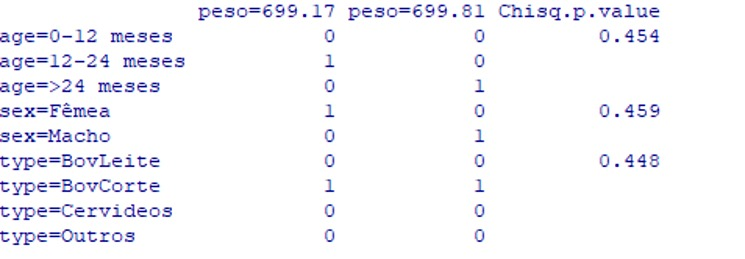
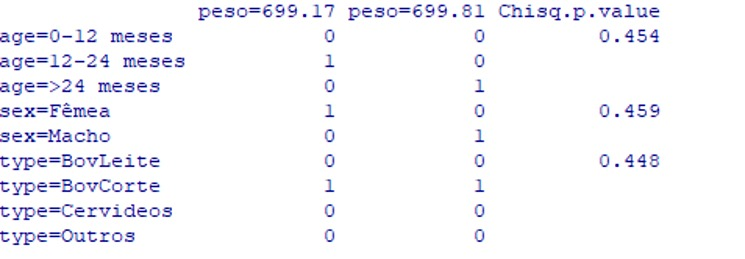
parasitas=0 87 0 **9.7-23**

O p-valor é menor que 0.05: rejeita-se H0, as variáveis apresentam associação, ou seja, a quantidade de parasitas pode estar relacionada com a ocorrência de tuberculose. Todos os animais doentes apresentavam alguma quantidade de parasitas, enquanto todos animais sadios não tinham nenhuma quantidade de parasitas.

1. **Se possível, demostre graficamente tais possíveis associações;**

Não é possível demonstrar graficamente tais associações.

1. **Também gostaria de saber se o peso estaria associado com o tipo do animal, o sexo, a idade e ao número de parasitas.**



**PESO X TIPO DE ANIMAL**

Pearson's Chi-squared test

data: .Table

X-squared = 393, df = 390, **p-value = 0.4479**

****

**PESO X SEXO**

Pearson's Chi-squared test

data: .Table

X-squared = 131, df = 130, **p-value = 0.4589**

****

**PESO X IDADE**

Pearson's Chi-squared test

data: .Table

X-squared = 262, df = 260, **p-value = 0.4535**

****

Os valores de p-valor de tipo, sexo e idade do animal são maiores que 0.05, sendo assim, aceita-se H0, verificamos que essas variáveis não apresentam associação com peso.

**PESO X QTD PARASITAS**

> #####Test for Pearson's correlation#####

> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0,1,0))

> scatterplot(parasitas~peso, regLine=list(method=lm, lty=1), smooth=FALSE, boxplots='xy',

+ data=tbrealprova)

> res <- NULL

> (res <- cor.test(tbrealprova$parasitas, tbrealprova$peso, alternative="two.sided",

+ method="pearson"))

**Pearson's product-moment correlation**

data: tbrealprova$parasitas and tbrealprova$peso

t = -0.59531, df = 129, **p-value = 0.5527**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval: -0.2218756 0.1202631

sample estimates: cor = -0.05234211

correlation coefficient = -0.0523, 95% CI -0.222-0.12, **p.value = 0.553**

Como r = -0.0523 e o p − valor = 0.553, aceita H0, verificamos que as variáveis peso e quantidade de parasitas também não apresentam correlação.



CONCLUSÃO: O objetivo deste trabalho é investigar os efeitos da idade, sexo, peso e tipo de animal no acometimento da doença. Após análise estatística, conclui-se que a idade apresenta efeito no acometimento da tuberculose: animais com mais de 24 meses apresentam maior ocorrência de tuberculose. As variáveis sexo, peso e tipo de animal não demonstraram efeito no acometimento da doença. Entretanto, em relação à quantidade de parasitas no animal, esta pode estar relacionada com a ocorrência de tuberculose, pois todos os animais doentes apresentavam alguma quantidade de parasitas, enquanto os animais sadios não tinham nenhuma quantidade de parasitas.