Universidade Federal Rural do Rio de Janeiro Pós-Graduação em Ciências Veterinárias

Avaliação da Disciplina de Métodos Estatísticos Aplicados às Ciências Veterinárias - 2020 Prof. Wagner Tassinari

Aluno(a): **Adriani da S. Carneiro Lopes e Laura N. Meirelles** Matrícula: **20191002974**/**20191002992**

**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

Peso: Quantitativa contínua;

Número de parasitas: Quantitativa discreta;

Sexo: Qualitativa nominal;

Idade: Qualitativa nominal;

Tipo de animal:qualitativa nominal;

Tuberculose: Qualitativa nominal;

Número do animal: quantitativa discreta;

Identificação da fazenda: quantitativa discreta.

* Análise exploratória ou descritiva:

1. Transformar para fator e colocar os devidos *labels* das variáveis categóricas;
2. Fazer o sumario estatístico das variáveis quantitativas e as distribuições de frequências para as variáveis qualitativas;
3. Plotar os devidos gráficos das variáveis para mostrar as distribuições de cada uma delas;

* Inferência Estatística:

1. Verificar se a patologia está associada ao: tipo do animal, sexo, idade, peso e a quantidade de parasitas;
2. Se possível, demostre graficamente tais possíveis associações;
3. Também gostaria de saber se o peso estaria associado com o tipo do animal, o sexo, a idade e ao número de parasitas.

**Obs 1:** Os resultados e as interpretações das análises deverão ser entregues em um arquivo no formato de um relatório do tipo .docx.

**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.

**Resolução:**

i) Transformar para fator e colocar os devidos *labels* das variáveis categóricas:

> Prova <-

+ readXL("C:/Users/adria/OneDrive/Documentos/Mestrado/Métodos Estatísticos/Dados/tbreal\_prova2020.xlsx",

+ rownames=FALSE, header=TRUE, na="", sheet="dados", stringsAsFactors=TRUE)

> Prova <- within(Prova, {

+ sex <- factor(sex, labels=c('Fêmea','Macho'))

+ })

> Prova <- within(Prova, {

+ type <- factor(type, labels=c('Bov Leite','Bov Corte','Cerv','Outros'))

+ })

> Prova <- within(Prova, {

+ age <- factor(age, labels=c('0-12 meses','12-24 meses','> 24 meses'))

+ })

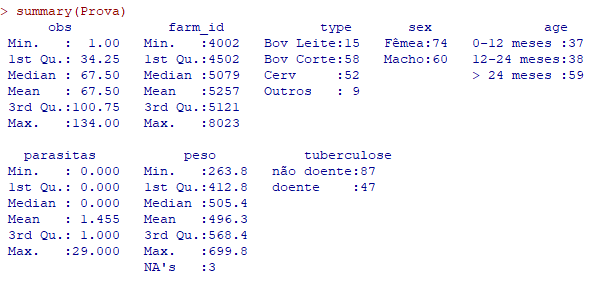
> Prova <- within(Prova, {

+ tuberculose <- factor(tuberculose, labels=c('não doente','doente'))

+ })

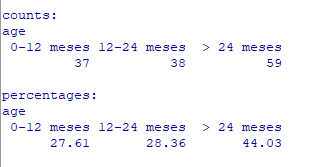
ii) Fazer o sumario estatístico das variáveis quantitativas e as distribuições de frequências para as variáveis qualitativas:

**Variáveis quantativas**:

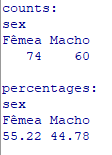


**Variáveis Qualitativas:**

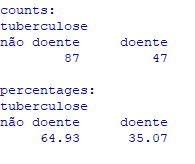
**Idade**

****

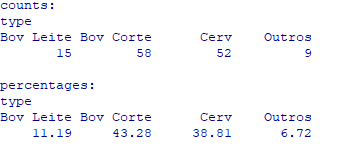
**Sexo**

****

**Tuberculose**

****

**Tipo**

****

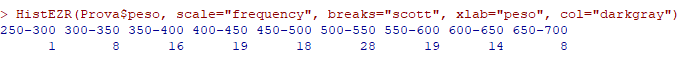
iii) Plotar os devidos gráficos das variáveis para mostrar as distribuições de cada uma delas.

**Variáveis quantitativa:**

Observação: Optou-se por utilizar os histograma porque o peso é uma variável quantitiva contínua.

**Peso**





**Parasitas:**

****

****

**Variáveis qualitativas:**

Observação: Optou-se por utilizar gráfico de barras ou colunas por se tratarem de variáveis qualitativas.

**Idade**

****

**Sexo**

****

**Tuberculose**

****

**Tipo**

****

**Inferência Estatística:**

**Normalidade dos dados**

A normalidade dos dados foi verificada através do teste de Shapiro-Wilk.

**PESO-TUBERCULOSE**

tuberculose = doente

Shapiro-Wilk normality test

data: peso

W = 0.97824, p-value = 0.5216

--------

p-values adjusted by the Holm method:

unadjusted adjusted

não doente 0.15526 0.31051

doente 0.52164 0.52164

**Resposta:** Como p-valor>0.05, acaita-se a hipótese que a distribuição estudada é normal.

**PESO-TIPO DO ANIMAL**

type = Bov Leite

Shapiro-Wilk normality test

data: peso

W = 0.93159, p-value = 0.2882

--------

type = Bov Corte

Shapiro-Wilk normality test

data: peso

W = 0.98333, p-value = 0.6177

--------

type = Cerv

Shapiro-Wilk normality test

data: peso

W = 0.98705, p-value = 0.8547

--------

type = Outros

Shapiro-Wilk normality test

data: peso

W = 0.9697, p-value = 0.8921

--------

p-values adjusted by the Holm method:

unadjusted adjusted

Bov Leite 0.28816 1

Bov Corte 0.61771 1

Cerv 0.85473 1

Outros 0.89211 1

**Resposta:** Como p-valor>0.05, aceita-se a hipótese que a distribuição estudada é normal.

**PESO-SEXO**

sex = Fêmea

Shapiro-Wilk normality test

data: peso

W = 0.97836, p-value = 0.2583

--------

sex = Macho

Shapiro-Wilk normality test

data: peso

W = 0.98234, p-value = 0.5351

--------

p-values adjusted by the Holm method:

unadjusted adjusted

Fêmea 0.25826 0.51652

Macho 0.53509 0.53509

**Resposta:** Como p-valor>0.05, aceita-se a hipótese que a distribuição estudada é normal.

**PESO-IDADE**

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

--------

age = 0-12 meses

Shapiro-Wilk normality test

data: peso

W = 0.96769, p-value = 0.3658

**Resposta:** Como p-valor>0.05, aceita-se a hipótese que a distribuição estudada é normal.

**PESO**

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

Shapiro-Wilk normality test

data: peso

W = 0.98562, p-value = 0.1847

tuberculose = não doente

Shapiro-Wilk normality test

data: peso

W = 0.97779, p-value = 0.1553

**Resposta:** Como p-valor>0.05, aceita-se a hipótese que a distribuição estudada é normal.

**NÚMERO DE PARASITAS**

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

Shapiro-Wilk normality test

data: parasitas

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

> Prova$parasitasexp <- with(Prova, exp(parasitas))

> normalityTest(~parasitasexp, test="shapiro.test", data=Prova)

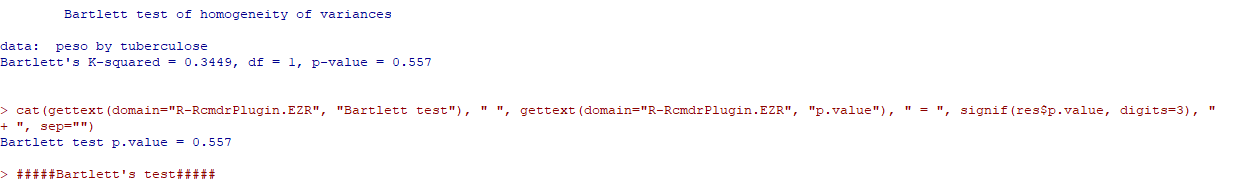
Shapiro-Wilk normality test

data: parasitasexp

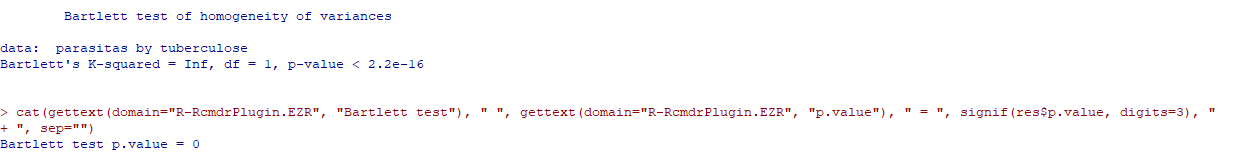
W = 0.060158, p-value < 2.2e-16

**Resposta:** Como p-valor<0.05, rejeita-se a hipótese que a distribuição estudada é normal.

**Teste para comparação entre as variâncias populacionais (Bartlett test)**

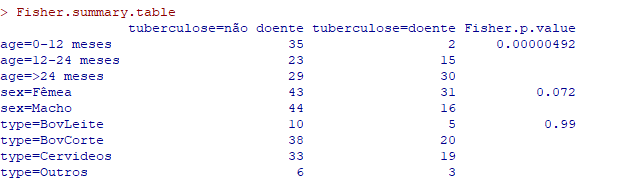


**Resposta:** Como p-valor=0.557, não rejeita-se H0. Sendo assim, a variabilidade do peso entre animais doentes e não doentes é homogênea.

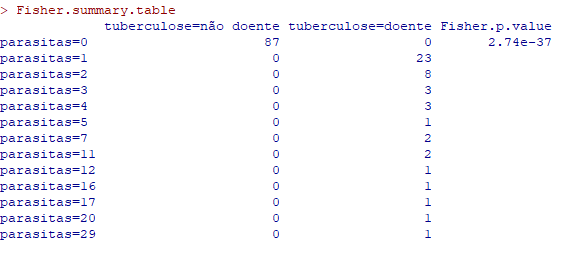


**Resposta:** Como p-valor<2,2.10-16 rejeita-se H0. Sendo assim, a variabilidade do número de parasitas entre animais doentes e não doentes é heterogênea.

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



Observação: Como no Bartlett test foi verificada que a variabilidade do número de parasitas entre animais doentes e não doentes é heterogênea, foi marcada a opção de correção de variável antes que o teste de Fisher fosse rodado.



**Respostas:**

Como o p-valor=0.00000492, rejeita-se H0 e portanto a patologia está associada com a idade do animal.

Como o p-valor=0.072, não rejeita-se H0 e portanto a patologia não está associada com sexo do animal.

Como o p-valor=0.99, não rejeita-se H0 e portanto a patologia não está associada com o tipo do animal.

Como o p-valor=2.74×10-37, rejeita-se H0 e portanto a patologia está associada ao número de parasitas.

Observação: Pelo fato da variável peso apresentar um grande número de amostras, foi possível rodar somente o teste do qui-quadrado.

> Fisher.summary.table

tuberculose=não doente tuberculose=doente Chisq.p.value

peso=263.75 0 1 0.459

peso=301.99 1 0

peso=310.52 1 0

peso=311.32 1 0

peso=311.85000000000002 1 0

peso=321.92 0 1

peso=324.23 1 0

peso=335.36 1 0

peso=341.82 1 0

peso=355.19 1 0

peso=357 0 1

peso=358.03 1 0

peso=359.32 1 0

peso=367.63 1 0

peso=372.7 1 0

peso=372.99 1 0

peso=373.59 1 0

peso=379.66 0 1

peso=389.93 1 0

peso=392.13 0 1

peso=392.28 1 0

peso=395.38 0 1

peso=395.59 0 1

peso=395.74 0 1

peso=399.4 0 1

peso=401.54 1 0

peso=403.07 0 1

peso=403.54 1 0

peso=408.48 0 1

peso=408.93 1 0

peso=409.08 1 0

peso=410.74 1 0

peso=412.75 1 0

peso=412.76 0 1

peso=412.89 0 1

peso=415.55 1 0

peso=416.38 1 0

peso=418.93 1 0

peso=425.82 0 1

peso=437.21 0 1

peso=441.41 0 1

peso=446.38 1 0

peso=447.27 1 0

peso=447.55 1 0

peso=451.53 1 0

peso=457.7 0 1

peso=460.17 1 0

peso=464.53 1 0

peso=470.88 1 0

peso=471.97 1 0

peso=474.86 1 0

peso=479.34 1 0

peso=479.83 0 1

peso=480.42 0 1

peso=482.7 1 0

peso=486.33 1 0

peso=487.1 0 1

peso=487.23 0 1

peso=487.4 1 0

peso=488.34 1 0

peso=493.49 1 0

peso=499.33 1 0

peso=500.22 0 1

peso=502.3 1 0

peso=502.83 0 1

peso=505.43 1 0

peso=506.33 0 1

peso=506.64 1 0

peso=509.09 1 0

peso=509.5 1 0

peso=510.65 0 1

peso=518.09 1 0

peso=518.16 0 1

peso=518.87 1 0

peso=519.13 1 0

peso=519.39 1 0

peso=519.6 1 0

peso=523.45000000000005 1 0

peso=525.85 1 0

peso=530.38 0 1

peso=532.69000000000005 1 0

peso=534.69000000000005 0 1

peso=534.91999999999996 0 1

peso=535.02 1 0

peso=537.46 1 0

peso=539.30999999999995 1 0

peso=541 1 0

peso=541.98 1 0

peso=543.76 0 1

peso=545.55 1 0

peso=555.53 1 0

peso=556.05999999999995 1 0

peso=559.19000000000005 0 1

peso=560.04999999999995 0 1

peso=562.28 1 0

peso=563.49 1 0

peso=565.36 1 0

peso=565.85 1 0

peso=571.01 1 0

peso=574.14 0 1

peso=575.05999999999995 0 1

peso=575.91 1 0

peso=576.79 0 1

peso=578.02 1 0

peso=579.66999999999996 1 0

peso=591.88 1 0

peso=597.41999999999996 1 0

peso=598.74 1 0

peso=599.01 0 1

peso=601.2 0 1

peso=601.46 0 1

peso=603.87 0 1

peso=604.27 1 0

peso=604.42 0 1

peso=605.11 0 1

peso=605.53 1 0

peso=609.46 1 0

peso=610.48 1 0

peso=616.98 0 1

peso=624.61 1 0

peso=626.83000000000004 1 0

peso=641.65 0 1

peso=645.08000000000004 1 0

peso=656.15 0 1

peso=659.85 1 0

peso=661 1 0

peso=664.09 1 0

peso=683.66 1 0

peso=689.62 0 1

peso=699.17 0 1

peso=699.81 0 1

> .Table <- xtabs(~tuberculose+peso, data=Prova)

> pairwise.prop2.test(.Table, p.adj="bonferroni", test.function=chisq.test)

Pairwise comparisons using chisq.test

data: .Table

não doente

doente 0.46

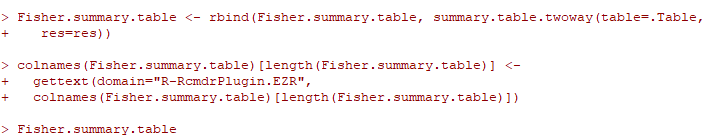
P value adjustment method: bonferroni

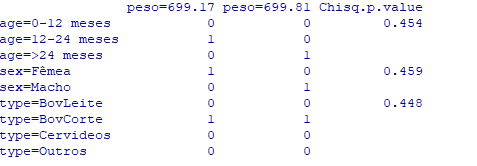
**Resposta:** Como o p-valor=0.459, não rejeita-se H0 e portanto a patologia não está associada com o peso do animal.

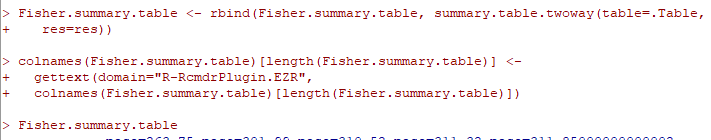
1. Se possível, demostre graficamente tais possíveis associações;

**Resposta:** Não é possível demonstrar graficamente, pois não há associações entre variáveis quantitativas.

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







**Respostas:**

Como o p-valor=0.454, não rejeita-se H0 e portanto o peso não está associada com a idade do animal.

Como o p-valor=0.459, não rejeita-se H0 e portanto o peso não está associada com o sexo do animal.

Como o p-valor=0.448, não rejeita-se H0 e portanto o peso não está associada com o tipo do animal.

Observação: para verificação da associação entre as variáveis quantitativas peso e número de parasitas foi escolhido o teste de Spearman, visto que a segunda variável apresenta uma distribuição não paramétrica.

Spearman's rank correlation rho

data: prova$parasitas and prova$peso

S = 355925, p-value = 0.5706

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.05000504

> cat(gettext(domain="R-RcmdrPlugin.EZR", "Spearman's rank correlation coefficient"),

+ signif(res$estimate, digits=3), gettext(domain="R-RcmdrPlugin.EZR", "p.value"), " = ",

+ signif(res$p.value, digits=3), "

+ ")

Spearman's rank correlation coefficient 0.05 p.value = 0.571



**Resposta:** Como o p-valor=0.571, não rejeita-se H0 e portanto o peso não está associada ao número de parasitas.