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): Amanda R. da C. Corval e Thaís Almeida Corrêa

**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.

**Tipos de Variáveis:**

Idade – Qualitativa Nominal, devido à categorização

Sexo – Qualitativa Nominal

Peso – Quantitativa Contínua

Tipo do Animal – Qualitativa Nominal

Parasitas – Quantitativa Discreta

Doença – Qualitativa Nominal

Identificação da Fazenda – Qualitativa Nominal

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

#####Convert numeric variables to factors#####

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

> #####Convert numeric variables to factors#####

> prova$type <- factor(prova$type, labels=c('Bovino Leiteiro','Bovino de Corte',

+ 'Cervídeos','Outros'))

> #####Convert numeric variables to factors#####

> prova$sex <- factor(prova$sex, labels=c('Fêmea','Macho'))

#####Convert numeric variables to factors#####

> prova$tuberculose <- factor(prova$tuberculose, labels=c('Não','Sim'))

> library(abind, pos=17)

> library(e1071, pos=18)

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

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

**+ "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))**

**mean sd IQR 0% 25% 50% 75% 100% n NA**

**parasitas 1.455224 4.012416 1.000 0.00 0.000 0.00 1.00 29.00 134 0**

**peso 496.296183 98.363228 155.675 263.75 412.755 505.43 568.43 699.81 131 3**

**> local({**

**+ .Table <- with(prova, table(age))**

**+ cat("\ncounts:\n")**

**+ print(.Table)**

**+ cat("\npercentages:\n")**

**+ print(round(100\*.Table/sum(.Table), 2))**

**+ })**

**- Idade:**

**counts:**

**age**

**0-12 meses 12-24 meses >24 meses**

**37 38 59**

**percentages:**

**age**

**0-12 meses 12-24 meses >24 meses**

**27.61 28.36 44.03**

**> local({**

**+ .Table <- with(prova, table(sex))**

**+ cat("\ncounts:\n")**

**+ print(.Table)**

**+ cat("\npercentages:\n")**

**+ print(round(100\*.Table/sum(.Table), 2))**

**+ })**

**-Sexo:**

**counts:**

**sex**

**fêmea macho**

**74 60**

**percentages:**

**sex**

**fêmea macho**

**55.22 44.78**

**> local({**

**+ .Table <- with(prova, table(tuberculose))**

**+ cat("\ncounts:\n")**

**+ print(.Table)**

**+ cat("\npercentages:\n")**

**+ print(round(100\*.Table/sum(.Table), 2))**

**+ })**

**- Tuberculose:**

**counts:**

**tuberculose**

**Não Sim**

**87 47**

**percentages:**

**tuberculose**

**Não Sim**

**64.93 35.07**

**> local({**

**+ .Table <- with(prova, table(type))**

**+ cat("\ncounts:\n")**

**+ print(.Table)**

**+ cat("\npercentages:\n")**

**+ print(round(100\*.Table/sum(.Table), 2))**

**+ })**

**-Tipo:**

**counts:**

**type**

**Bovino leiteiro Bovino de corte Cervídeos Outros**

**15 58 52 9**

**percentages:**

**type**

**Bovino leiteiro Bovino de corte Cervídeos Outros**

**11.19 43.28 38.81 6.72**

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

* **Para variáveis quantitativas – gráfico escolhido: Histograma**

****

* **Para variáveis qualitativas – gráfico escolhido: Barras**

** **

** **

**- Inferência Estatística:**

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

***Normalidade***

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

**Peso em relação ao 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

Sendo p-valor>0.05, aceita-se a hipótese que a variável estudada é normal.

**Peso em relação ao sexo do animal:**

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

Sendo p-valor>0.05, aceita-se a hipótese que a variável estudada é normal.

**Peso em relação a 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

Sendo p-valor>0.05, aceita-se a hipótese que a variável estudada é normal.

**Peso em relação a idade do animal:**

> 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

Sendo p-valor>0.05, aceita-se a hipótese que a variável 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

Sendo p-valor>0.05, aceita-se a hipótese que a variável 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

Sendo p-valor<0.05, rejeita-se a hipótese que a variável estudada é normal.

* Teste de Homogeneidade (Bartlett’s test)

Teste de avaliação da homogeneidade variâncias populacionais

Bartlett test of homogeneity of variances

data: parasitas by tuberculose

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

Devido ao p-valor ser <2,2.10-16 , rejeita-se H0. Sendo assim, a variabilidade do número de parasitas entre animais doentes e não doentes é heterogênea.

Bartlett test of homogeneity of variances

data: peso by tuberculose

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

Há homogeneidade dos dados

Devido ao p-valor ser igual a 0.557, não rejeitou-se H0. Sendo assim a variabilidade do peso dos animais doentes e não doentes apresenta-se de forma homogênea

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

* Fisher.summary.table
* tuberculose=não tuberculose=sim Fisher.p.value
* age=0-12 meses 35 2 0.00000492
* age=12-24 meses 23 15
* age=>24 meses 29 30
* sex=fêmea 43 31 0.072
* sex=macho 44 16
* type=Bovino leiteiro 10 5 0.99
* type=Bovino de Corte 38 20
* type=Cervídeos 33 19
* type=Outros 6 3

Fisher.summary.table

tuberculose=não tuberculose=sim Fisher.p.value

parasitas=0 87 0 2.74e-37

parasitas=1 0 23

parasitas=2 0 8

parasitas=3 0 3

parasitas=4 0 3

parasitas=5 0 1

parasitas=7 0 2

parasitas=11 0 2

parasitas=12 0 1

parasitas=16 0 1

parasitas=17 0 1

parasitas=20 0 1

parasitas=29 0 1

* 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.

Fisher.summary.table

tuberculose=não tuberculose=sim 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=311.85 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=523.45 1 0

peso=525.85 1 0

peso=530.38 0 1

peso=532.69 1 0

peso=534.69 0 1

peso=534.92 0 1

peso=535.02 1 0

peso=537.46 1 0

peso=539.31 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=556.06 1 0

peso=559.19 0 1

peso=560.05 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=575.06 0 1

peso=575.91 1 0

peso=576.79 0 1

peso=578.02 1 0

peso=579.67 1 0

peso=591.88 1 0

peso=597.42 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=626.83 1 0

peso=641.65 0 1

peso=645.08 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

sim 0.46

P value adjustment method: bonferroni

P-valor (0.459) não rejeita H0, portanto, não há associação entre as variáveis.

Associação entre tuberculose e parasitas

Fisher.summary.table

tuberculose=não tuberculose=sim Chisq.p.value

parasitas=0 87 0 9.7e-23

parasitas=1 0 23

parasitas=2 0 8

parasitas=3 0 3

parasitas=4 0 3

parasitas=5 0 1

parasitas=7 0 2

parasitas=11 0 2

parasitas=12 0 1

parasitas=16 0 1

parasitas=17 0 1

parasitas=20 0 1

parasitas=29 0 1

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

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

Pairwise comparisons using chisq.test

data: .Table

não

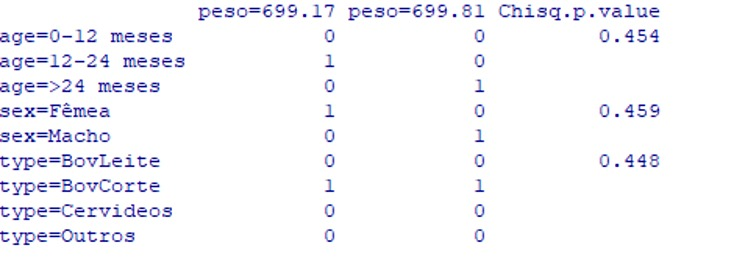
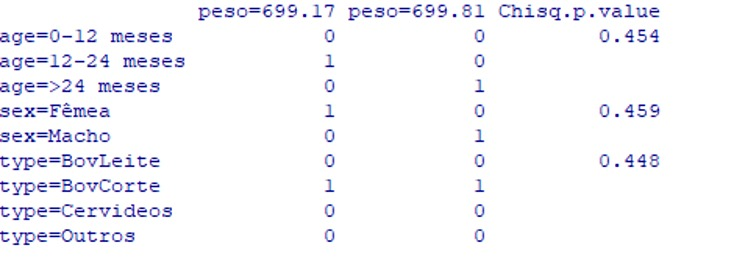
sim <2e-16

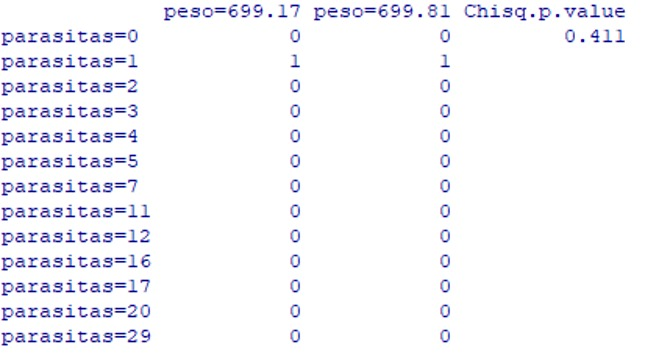
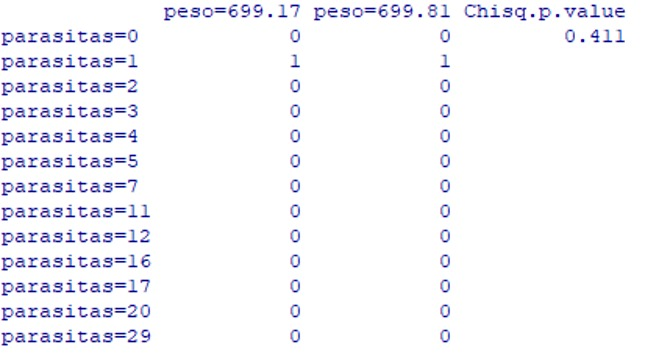
P value adjustment method: bonferroni

Como só há duas variáveis quantitativas, o único gráfico possível de se fazer é o da associação entre parasitas e peso.

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

Fisher.summary.table <- rbind(Fisher.summary.table, summary.table.twoway(table=.Table, res=res))





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

Não há correlação entre as variáveis analisadas devido ao p –valor ser maior que 0,05. Desta forma não rejeitamos H0 e confirmamos a falta de correlação entre as variáveis analisadas.

