Dias 1 e 2 - Curso R

Enggel Carmo

18/07/2021

# Curso R – Guia Prático para análise de variáveis categóricas

* Por que eu estou aprendendo isso?
  + Lembrar da Instalação do R
  + Instalação do R Studio
  + Instalação do pacote R Markdown
  + Abrir o R studio
  + Abrir o R markdown

# Dia 1- Criando um documento no R Markdown

* Será que as funções estão sendo salvas?
* Tem que dar espaço para a palavra :)
* Dar **dois espaços** no fim das linhas em branco para mostrar (ao programa) que estamos pulando uma linha (em negrito - entre asterisco)
* Para *itálico* (apenas um asteriscos)
* O R2 está sobrescrito (entre circunflexo)
* O teste12 está subescrito (entre til)
* Coisas com cara de código fica shapiro.wilk (entre acento grave)

Link clicável abaixo

[Meu linkedin:](https://www.linkedin.com/in/enggel-carmo/)

Adicionar imagem



### Criando referências

De acordo com Crawley 2012 [[1]](#footnote-25).

## Parte 1.

### Criar pasta para o curso

### Criar documento RMarkdown

### Salvar documento Dados\_CursoR.txt na pasta para o curso

### Salvar documento **Untitled** na pasta do curso

### Salvar documento .docx > Mostrar documento salvo

# Dia 2 - Importar dados - Distribuição Normal

## Parte 2. Instalar pacotes

### Instalar pacotes que serão necessários para o curso

library (MASS)  
library (car)  
library(multcompView)#TukeyHSD  
library(dplyr)  
library(rstatix)  
library(emmeans)  
library(ggplot2)  
library(gplots)

### Importar dados Dados\_CursoR.txt

Os formatos mais comuns para importar dados da sua máquina são .csv ou .txt

* Lembrar que .csv deve especificar que as colunas estão separadas por ;
* Anotações dentro e fora do *{r}*

dados<-read.table("Dados\_CursoR.txt",h=T) #comentários > h=T quer dizer que existe um cabeçalho  
dados

## Dieta rep ovos larvas survive censor total  
## 1 pulgao cr1 94 76 30 0 1  
## 2 pulgao cr2 73 50 30 0 1  
## 3 pulgao cr3 68 58 20 1 1  
## 4 pulgao cr4 96 72 30 0 1  
## 5 pulgao cr5 67 40 22 1 1  
## 6 pulgao cr7 80 46 20 1 1  
## 7 pulgao cr8 54 40 30 0 1  
## 8 pulgao cr9 94 77 15 1 1  
## 9 pulgao cr10 72 46 10 1 1  
## 10 pulgao cr11 88 67 22 1 1  
## 11 pulgao cr12 98 86 30 0 1  
## 12 pulgao cr13 94 70 30 0 1  
## 13 pulgao cr14 75 46 20 1 1  
## 14 pulgao cr15 56 40 30 0 1  
## 15 pulgao cr16 94 77 15 1 1  
## 16 cochonilha tt1 76 33 18 1 1  
## 17 cochonilha tt2 78 16 30 0 1  
## 18 cochonilha tt3 83 35 17 1 1  
## 19 cochonilha tt4 42 13 30 0 1  
## 20 cochonilha tt5 68 24 6 1 1  
## 21 cochonilha tt6 87 37 25 1 1  
## 22 cochonilha tt7 51 18 18 1 1  
## 23 cochonilha tt8 62 16 16 1 1  
## 24 cochonilha tt10 92 43 20 1 1  
## 25 cochonilha tt11 77 26 12 1 1  
## 26 cochonilha tt12 70 24 25 1 1  
## 27 cochonilha tt13 48 16 18 1 1  
## 28 cochonilha tt14 91 37 25 1 1  
## 29 cochonilha tt16 53 10 18 1 1  
## 30 cochonilha tt17 78 23 16 1 1  
## 31 ovos dz2 15 2 30 0 1  
## 32 ovos dz3 12 0 17 1 1  
## 33 ovos dz4 28 4 30 0 1  
## 34 ovos dz5 32 2 30 0 1  
## 35 ovos dz6 17 3 30 0 1  
## 36 ovos dz7 31 2 30 0 1  
## 37 ovos dz8 28 2 30 0 1  
## 38 ovos dz9 22 0 13 1 1  
## 39 ovos dz10 20 0 26 1 1  
## 40 ovos dz11 26 3 21 1 1  
## 41 ovos dz12 14 2 30 0 1  
## 42 ovos dz13 12 3 30 0 1  
## 43 ovos dz16 17 3 30 0 1  
## 44 ovos dz17 33 2 30 0 1  
## 45 ovos dz18 29 2 30 0 1

attach(dados) #direciona que os dados utilizados será do arquivo que estamos utilizando no momento fixando na análise  
summary(dados)

## Dieta rep ovos larvas   
## Length:45 Length:45 Min. :12.00 Min. : 0.00   
## Class :character Class :character 1st Qu.:29.00 1st Qu.: 3.00   
## Mode :character Mode :character Median :67.00 Median :24.00   
## Mean :57.67 Mean :28.71   
## 3rd Qu.:80.00 3rd Qu.:46.00   
## Max. :98.00 Max. :86.00   
## survive censor total   
## Min. : 6.00 Min. :0.0000 Min. :1   
## 1st Qu.:18.00 1st Qu.:0.0000 1st Qu.:1   
## Median :25.00 Median :1.0000 Median :1   
## Mean :23.44 Mean :0.5556 Mean :1   
## 3rd Qu.:30.00 3rd Qu.:1.0000 3rd Qu.:1   
## Max. :30.00 Max. :1.0000 Max. :1

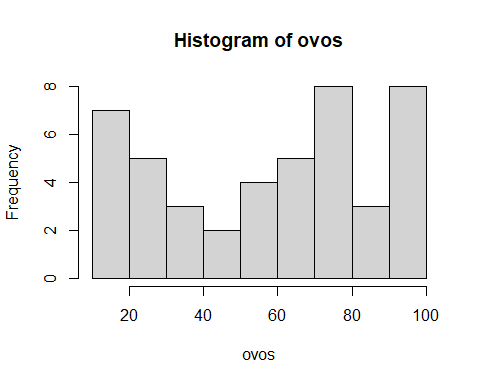
nrow(dados[which(dados$Dieta=="pulgao"),]) #tamanho do nível específico

## [1] 15

### Representação da variável **ovos**

Avaliaremos os dados de oviposição e sobrevivência da espécie *Eriopis connexa* em relação as **dietas** consumidas/ofertadas, logo são **variáveis categóricas**.

hist(ovos)

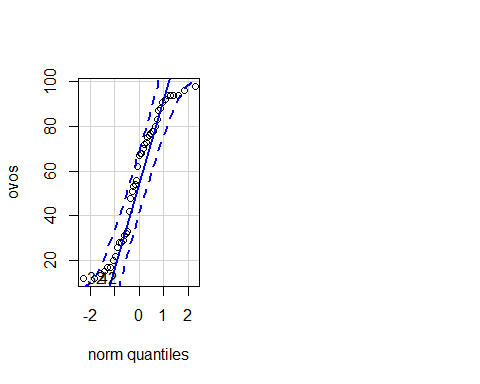


### Representação gráfica dos quantiles

Pode auxiliar na escolha do modelo que a distribuição dos seus dados melhor se encaixa.

par(mfrow=c(1,2)) # plot de dois gráficos na mesma folha  
qqp(ovos, "norm") #quantile quantile plot da distribuição normal

## [1] 32 42



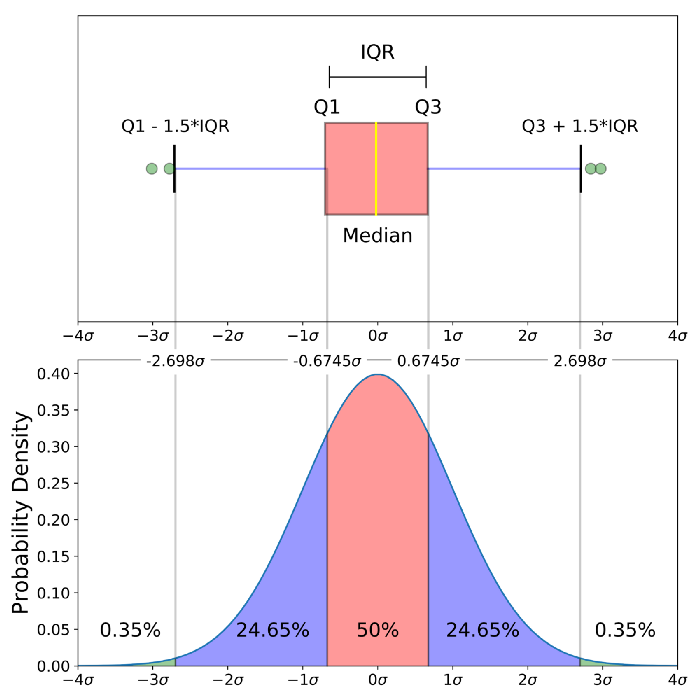
## Parte 3. Análise para dados com distribuição Normal ou Gaussiana

### Vamos testar a distribuição para a variável **ovos**

* Distribuição Normal ou Gaussiana
  + Concentração de valores em torno de um valor central;
  + Simetria em torno do valor central;
  + Frequência pequena de valores muito extremos.
* A distribuição normal possui dois parâmetros, a média (μ), ou seja onde está centralizada e a variância (σ2>0) que descreve o seu grau de dispersão. Ou ainda o desvio padrão (σ). Cabe salientar que como qualquer outro modelo, dependendo dos parâmetros, teremos diferentes distribuições normais.

Aceitar  
H0: consistem em dizer que há normalidade nos dados - p > 0.05

H1: que não há - p < 0.05



shapiro.test(ovos)

##   
## Shapiro-Wilk normality test  
##   
## data: ovos  
## W = 0.90943, p-value = 0.001888

### Vamos testar a homocedasticidade para **ovos**

Homocedasticidade é o termo para designar variância constante dos erros experimentais. *Cada nível tem sua variação em torno da média, se estes apresentam grande variação, mostra que os dados não são* ***homogêneos****.*

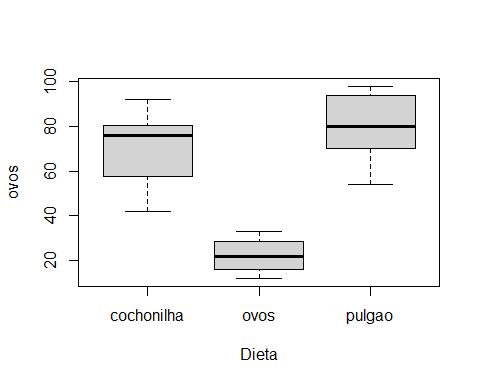
Quando *p-valor* for **menor** que o nível de significância escolhido para o teste (em geral, convencionado em 0,05), rejeita-se a H0 e aceitamos hipótese alternativa, assumindo a **desigualdade das variâncias**. Caso ele seja **maior** que o nível de significância, concluímos que **as variâncias são constantes ou homogêneas**.

bartlett.test(ovos ~ Dieta, data = dados)

##   
## Bartlett test of homogeneity of variances  
##   
## data: ovos by Dieta  
## Bartlett's K-squared = 7.5519, df = 2, p-value = 0.02291

### Boxplot

boxplot(ovos~Dieta)



### *ANOVA*

modelo1<-aov(ovos~Dieta)  
anova(modelo1,test="F")

## Analysis of Variance Table  
##   
## Response: ovos  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Dieta 2 28704.4 14352.2 80.227 4.523e-15 \*\*\*  
## Residuals 42 7513.6 178.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Agora podemos fazer o teste t para separação de médias

Tukey<-TukeyHSD(x=modelo1,'Dieta', conf.level = 0.95)  
Tukey

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = ovos ~ Dieta)  
##   
## $Dieta  
## diff lwr upr p adj  
## ovos-cochonilha -48.0 -59.865459 -36.13454 0.0000000  
## pulgao-cochonilha 9.8 -2.065459 21.66546 0.1230937  
## pulgao-ovos 57.8 45.934541 69.66546 0.0000000

Extra: Shapiro-Wilk de cada nível dentro do grupo

library(RVAideMemoire)  
byf.shapiro(ovos~Dieta, data=dados)

##   
## Shapiro-Wilk normality tests  
##   
## data: ovos by Dieta   
##   
## W p-value   
## cochonilha 0.9375 0.35216   
## ovos 0.9083 0.12737   
## pulgao 0.8939 0.07693 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Parte 4. Como representar essas informações no trabalho?

#### Nos resultados: O tipo de dieta apresentou efeito ou diferença significativa sobre a fecundidade de *Eriopis connexa* (F2,3319.7, P<0.001).

## Análise descritiva

### Médias e erros

medias<-tapply(ovos, Dieta, mean)  
medias

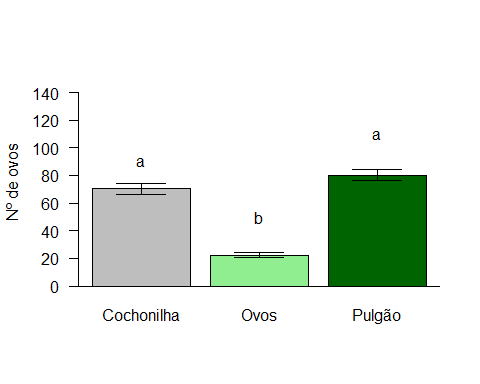
## cochonilha ovos pulgao   
## 70.4 22.4 80.2

erro<-tapply(ovos, Dieta, sd)/sqrt(tapply(ovos, Dieta, length))  
erro

## cochonilha ovos pulgao   
## 4.128646 1.956187 3.860915

### Representação em Barplot

barplot2(medias,  
 plot.ci=T,  
 ci.u=medias+erro,  
 ci.l=medias-erro,  
 ylab="Nº de ovos",  
 ylim=c(0,150),  
 las=1,  
 names.arg = c("Cochonilha","Ovos", "Pulgão"),  
 col=c("gray","lightgreen", "darkgreen"))  
 text(0.7,90,"a")  
 text(1.9,50,"b")  
 text(3.1,110,"a")  
abline(h=0)



Livro R markdown [[2]](#footnote-54). [Livro:](https://bookdown.org/yihui/bookdown/)

**Lembrar de fazer o detach para desafixar a tabela do programa.**

detach(dados)

**Risco de “chamar” variável de outra tabela por ter nomes idênticos.**

**Salvar documento .docx > Mostrar documento salvo**

#### Exercício: Façam o mesmo script para larvas.

Quem souber transformar para Box-Cox, log ou sqrt fique à vontade também ou ver se a distribuição se encaixa melhor para Gamma.

Vem aí:

# Dia 3 - E quando nossos dados não apresentam distribuição normal?

* GLM - Distribuição Poisson
* GLM - Distribuição Binomial Negativa.
* Análise de contraste.

1. Crawley, M. J. (2012). The R book. John Wiley & Sons. [↑](#footnote-ref-25)
2. Xie, Y. (2016). Bookdown: authoring books and technical documents with R markdown. CRC Press. [↑](#footnote-ref-54)