Trabalho Final de Inferência Estatística

Carlos Alberto Torres Quintanilha Neto

Turma: MRJ02021-TBABD-8

18/09/2020

##### 

# Carregando o Datasets  
# Dataset da Questão 1  
dadosq1 <- read.csv("dados\_tratamento\_covid.csv", header = TRUE,sep = ";", dec = ",")  
# Dataset da Questão 2  
dadosq2 <- read.csv("tempo\_bateria.csv", header = TRUE,sep = ";", dec = ",")  
# Dataset da Questão 3  
dadosq3 <- read.csv("qualidade\_restaurante.csv", header = TRUE,sep = ";", dec =",")

# Questão 1

Uma fabricante da indústria farmacêutica está em fase de pesquisa e desenvolvimento de um medicamento para combate ao COVID-19. Foi selecionada uma amostra de 630 pessoas contaminadas para fins de testes iniciais do tratamento no Brasil, apresentadas no banco de dados “dados\_tratamento\_covid.csv”. Para cada uma destas pessoa, foram avaliadas o resultado do tratamento (cura ou óbito) e a ocorrência de efeitos colaterais (sim ou não).

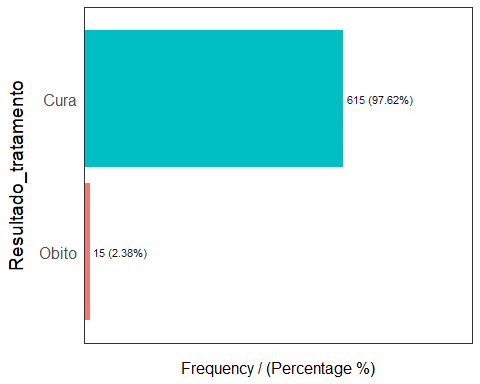
## Entendento o DataSet

# Head do dataset para uma flextable  
pander::pander((head(dadosq1)))

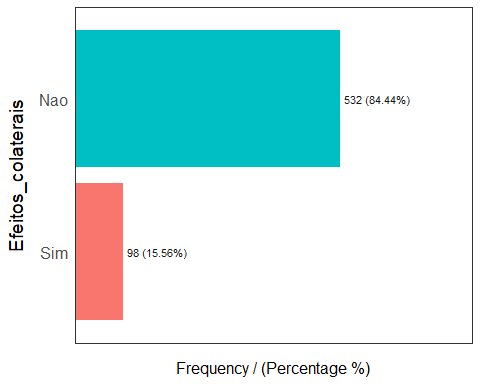
|  |  |  |
| --- | --- | --- |
| ID\_paciente | Resultado\_tratamento | Efeitos\_colaterais |
| 1 | Cura | Nao |
| 2 | Cura | Nao |
| 3 | Cura | Nao |
| 4 | Cura | Nao |
| 5 | Cura | Nao |
| 6 | Cura | Nao |

##### 

# Apresentação de informações gerais sobre as variaveis do Dataset.  
freq(dadosq1)



## Resultado\_tratamento frequency percentage cumulative\_perc  
## 1 Cura 615 97.62 97.62  
## 2 Obito 15 2.38 100.00



## Efeitos\_colaterais frequency percentage cumulative\_perc  
## 1 Nao 532 84.44 84.44  
## 2 Sim 98 15.56 100.00

## [1] "Variables processed: Resultado\_tratamento, Efeitos\_colaterais"

## Primeira Pergunta da Questão 1 - Letra a)

Considerando que a taxa de letalidade geral da COVID- 19 é de 3,4%, podemos afirmar que este medicamento tem efetividade na redução da proporção de óbitos? Considere um nível de significância de 0,01 e realize as verificações necessárias a respeito das premissas para realização deste teste de hipóteses. Explique a saída encontrada no R e interprete os resultados.

**Resolução:**

Para responder o questionamento foi montada a seguinte Hipótese:

H0: p >= 3.4%  
 Ha: p < 3.4%

Será realizado o Teste Unicaldal inferior.

#Tamanho da Amostra (n)  
  
n <- nrow(dadosq1)  
n

## [1] 630

#Quantidade de óbitos da amostra  
obitos <- sum(dadosq1$Resultado\_tratamento == "Obito")  
obitos

## [1] 15

# Proporçao da Amostra (pbar)  
pbar <- obitos/n  
pbar

## [1] 0.02380952

# Valor da hipótese(p0)  
p0 <- 0.034  
  
# Calculando a estatística Z  
z = (pbar−p0)/sqrt(p0∗(1−p0)/n)  
z

## [1] -1.411357

# Computando o valor Crítico  
alpha= 0.01  
z.alpha = qnorm(1-alpha)  
-z.alpha

## [1] -2.326348

# Como z > z.alpha não é possível rejeitar H0  
  
# Solução alternativa 1 , calculando o p-value da (pval)  
pval = pnorm(z)  
pval

## [1] 0.07906976

# Como pval > 0.01 não é possível rejeitar H0  
  
# Solução alternativa 2, utilizando a função prop.test  
prop.test(obitos,n,p=0.034, alt="less", correct=FALSE, conf.level = 0.99)

##   
## 1-sample proportions test without continuity correction  
##   
## data: obitos out of n, null probability 0.034  
## X-squared = 1.9919, df = 1, p-value = 0.07907  
## alternative hypothesis: true p is less than 0.034  
## 99 percent confidence interval:  
## 0.00000000 0.04250806  
## sample estimates:  
## p   
## 0.02380952

# Como p-value > 0.01 não é possível rejeitar H0

As distribuições foram consideradas normais pois:

n.p => 5 e n.(1-p) => 5

As 3 soluções apontam para a não rejeição da hipótese nula.

* No primeiro caso z (-1.411357) é maior que o p-value crítico (-2,326348);
* No segundo caso o p-value da cauda inferior é (0.07906976), maior que 0.01;
* No terceiro caso p-value(0.07907) é maior que 0.01.

A não rejeição da hipótese nula significa que não é possível confirmar a efetividade do medicamento na redução no número de óbitos.

## Segunda Pergunta da Questão 1 - Letra b)

Outros medicamentos já testados no mundo apresentam, em média, efeitos colaterais em 18% dos pacientes. Podemos afirmar que este novo medicamento gera uma menor proporção de efeitos colaterais? Considere um nível de significância de 0,01 e realize as verificações necessárias a respeito das premissas para realização deste teste de hipóteses. Explique a saída encontrada no R e interprete os resultados.

Resolução:

Para responder o questionamento foi montada a seguinte Hipótese:

H0: p >= 0.18

Ha: p < 0.18

Será realizado o Teste Unicaldal inferior.

#Tamanho da Amostra (n)  
n <- nrow(dadosq1)  
n

## [1] 630

#Quantidade de pacientes com efeitos colaterais na amostra  
efeitos <- sum(dadosq1$Efeitos\_colaterais == "Sim")  
efeitos

## [1] 98

# Proporçao da Amostra (pbar)  
pbar <- efeitos/n  
pbar

## [1] 0.1555556

# Valor da hipótese(p0)  
p0 <- 0.18  
p0

## [1] 0.18

# Calculando a estatística Z  
z = (pbar−p0)/sqrt(p0∗(1−p0)/n)  
z

## [1] -1.597009

# Computando o valor Crítico  
alpha= 0.01  
z.alpha = qnorm(1-alpha)  
-z.alpha

## [1] -2.326348

# Como z > z.alpha não é possível rejeitar H0  
  
# Solução alternativa 1 , calculando o p-value da (pval)  
pval = pnorm(z)  
pval

## [1] 0.05513189

# Como pval e > 0.01 não é possível rejeitar H0  
  
# Solução alternativa 2, utilizando a função prop.test  
prop.test(efeitos,n,p=0.18, alt="less", correct=FALSE,conf.level = 0.99)

##   
## 1-sample proportions test without continuity correction  
##   
## data: efeitos out of n, null probability 0.18  
## X-squared = 2.5504, df = 1, p-value = 0.05513  
## alternative hypothesis: true p is less than 0.18  
## 99 percent confidence interval:  
## 0.000000 0.192066  
## sample estimates:  
## p   
## 0.1555556

# Como p-value > 0.01 não é possível rejeitar H0

As 3 soluções apontam para a não rejeição da hipótese nula.

* No primeiro caso z (-1.597009) é maior que o p-value crítico (-2,326348)
* No segundo caso o p-value da cauda inferior é (0.05513189), maior que 0.01
* No terceiro caso p-value(0.05513) é maior que 0.01

A não rejeição da hipótese nula significa que não é possível confirmar a efetividade do medicamento na redução do número de pacientes com efeitos colaterais.

# Questão 2

Uma fabricante de baterias de celular está testando uma nova tecnologia para aumentar o tempo de duração. Para isso, ela está realizando um experimente com o objetivo de comparar o tempo de duração de suas baterias atuais com o tempo de duração das baterias que utilizam a nova tecnologia (banco de dados “tempo\_bateria.csv”).

## Primeira Pergunta da Questão 2 - Letra a)

Realize análise exploratória dos dados e explicite as principais conclusões encontradas.

# Carregando o head do dataset  
pander::pander((head(dadosq2)))

|  |  |
| --- | --- |
| tempo\_atual | tempo\_nova |
| 17.3 | 20.6 |
| 13.9 | 22.5 |
| 18.8 | 12.3 |
| 24 | 21.3 |
| 20 | 17.6 |
| 30.1 | 9.6 |

# Apresentação de informações gerais sobre as variaveis do Dataset.  
  
# Detalhes sobre a amostras  
pander::pander(summary(dadosq2))

|  |  |
| --- | --- |
| tempo\_atual | tempo\_nova |
| Min. : 4.50 | Min. : 5.50 |
| 1st Qu.:11.50 | 1st Qu.:12.07 |
| Median :13.80 | Median :16.00 |
| Mean :14.85 | Mean :17.18 |
| 3rd Qu.:17.00 | 3rd Qu.:20.68 |
| Max. :34.20 | Max. :36.90 |
| NA | NA’s :65 |

nrow(dadosq2)

## [1] 155

# São 155 amostras sendo que há 65 NA´s nas amostras de tempo de duração com a nova tecnologia.  
  
# Invertendo e remontando o dataframe para facilitar o tratamento.  
dadosq2\_invertido <- dadosq2 %>%  
 rownames\_to\_column() %>%  
 pivot\_longer(-rowname,"Tecnologia", values\_to = "Tempo")  
mu <- dadosq2\_invertido %>%  
 group\_by(Tecnologia) %>%  
 summarise(media = mean(Tempo,na.rm = T))

## `summarise()` ungrouping output (override with `.groups` argument)

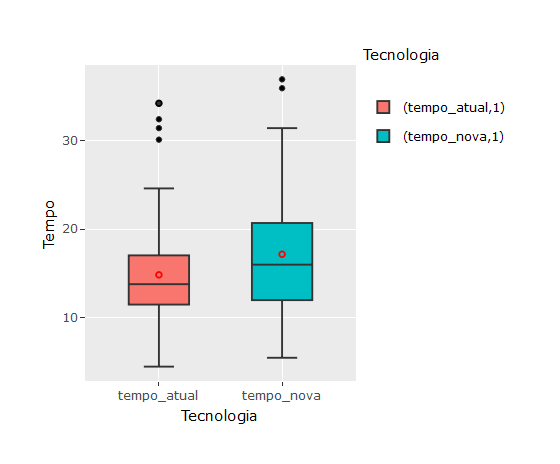
# Retirando os NA´s apenas das amostras da nova bateria.  
  
dadosq2\_invertido <- na.omit(dadosq2\_invertido)  
  
# Montando uma lista para facilitar o trabalho  
listatec <- list( tempo\_atual = dadosq2\_invertido$Tempo[dadosq2\_invertido$Tecnologia == "tempo\_atual"],  
 tempo\_nova = dadosq2\_invertido$Tempo[dadosq2\_invertido$Tecnologia == "tempo\_nova"])  
  
# Mais detalhes sobre as amostras  
#autofit(flextable(status(dadosq2)))  
pander::pander(status(dadosq2))

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | q\_zeros | p\_zeros | q\_na | p\_na | q\_inf | p\_inf | type | unique |
| tempo\_atual | 0 | 0 | 0 | 0 | 0 | 0 | numeric | 100 |
| tempo\_nova | 0 | 0 | 65 | 0.4194 | 0 | 0 | numeric | 73 |

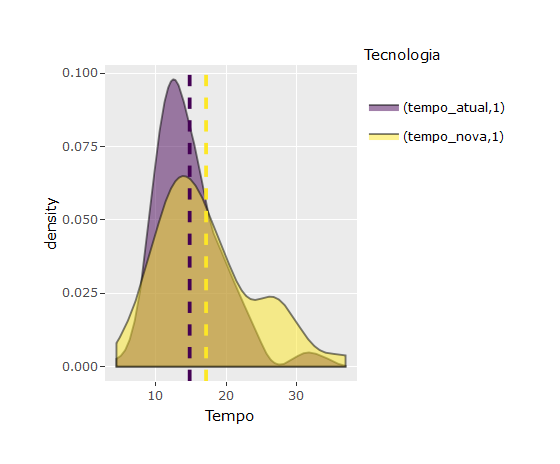
autofit(flextable(profiling\_num(dadosq2),col\_keys=c("variable","mean","std\_dev","skewness","kurtosis")))

| variable | mean | std\_dev | skewness | kurtosis |
| --- | --- | --- | --- | --- |
| tempo\_atual | 14.85484 | 4.891952 | 1.2250076 | 5.347449 |
| tempo\_nova | 17.17556 | 6.865650 | 0.7444421 | 3.043392 |

# Gráficos de caixa das amostras  
dadosq2\_invertido %>%  
 ggplot(aes(x=Tecnologia,y=Tempo, fill = Tecnologia)) %+%  
 geom\_boxplot(show.legend = F) %+%  
 scale\_color\_brewer(palette="Dark2") %+%   
 stat\_summary(fun=mean,geom="point",color="red")%>%  
 ggplotly()



# Gráficos de Densidade das Amostras  
  
dadosq2\_invertido %>%  
 ggplot(aes(x=Tempo, fill = Tecnologia)) %+%  
 geom\_density(alpha = 0.5) %+%  
 geom\_vline(data=mu,show.legend = F, aes(xintercept = media, color = Tecnologia), linetype = 2, size=1) %+%  
 scale\_fill\_viridis\_d() %+%  
 scale\_color\_viridis\_d() %>% ggplotly()



# Testes de Normalidade  
# Shapiro  
  
listatec %>%  
 map(~shapiro.test(.x))

## $tempo\_atual  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.92351, p-value = 2.476e-07  
##   
##   
## $tempo\_nova  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.94998, p-value = 0.00167

# kolmogorov-smirnov  
listatec %>%  
 map(~ks.test(.x,"pnorm",mean(.x),sd(.x)))

## Warning in ks.test(.x, "pnorm", mean(.x), sd(.x)): ties should not be present  
## for the Kolmogorov-Smirnov test  
  
## Warning in ks.test(.x, "pnorm", mean(.x), sd(.x)): ties should not be present  
## for the Kolmogorov-Smirnov test

## $tempo\_atual  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.095179, p-value = 0.1206  
## alternative hypothesis: two-sided  
##   
##   
## $tempo\_nova  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.1198, p-value = 0.151  
## alternative hypothesis: two-sided

Verifica-se que no teste Shapiro acaba não sendo possível rejeitar a hipótese de que ambas as distribuições não seguem a normal dado que o p-value é, nos dois casos, menor que 0.05. No teste kolmogorov-smirnov é possível rejeitar a hipótese e assumir que ambas as amostras seguem a distribuição normal. Esta diferença se deve, possivelmente, à sensibilidade do teste Shapiro ao tamanho da amostra. Pelos gráficos e pela medida do “Skewness” é possível identificar que há assimetria com calda a direita e que a amostra de tempo atual apresenta um kurtosis bem maior que 3 apresentando boa probabilidade de possuir outliers.

Como a amostra é maior que 30 seguiremos na resolução da questão considerando a normalidade das distribuições.

## Segunda Pergunta da Questão 2 - Letra b)

Há evidências de que a bateria com nova tecnologia gera aumento no tempo de duração quando se comparada com o atual modelo? Considere um nível de significância de 0,05 e realize as verificações necessárias a respeito das premissas para realização deste teste de hipóteses. Explique a saída encontrada no R e interprete os resultados. As hipóteses são:

* H0: 00 - 1 => 0
* Ha: 00 - 1 < 0

t.test(dadosq2$tempo\_atual,na.omit(dadosq2$tempo\_nova),mu=0,alternative = "less")

##   
## Welch Two Sample t-test  
##   
## data: dadosq2$tempo\_atual and na.omit(dadosq2$tempo\_nova)  
## t = -2.8181, df = 142.07, p-value = 0.00276  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -0.9573007  
## sample estimates:  
## mean of x mean of y   
## 14.85484 17.17556

Foi rodada a função Ttest, como o p-value(0.0.00276) é menor que 0.05, pode-se rejeitar a hipótese nula de que a diferença entre a média autal e a média nova seja igual ou maior que 0 apontando para um real aumento no tempo de duração.

# Questão 3

O gerente de uma rede de restaurantes deseja melhorar o nível de qualidade do atendimento em suas unidades. Para isso, ele realizou um programa intensivo de treinamento em 25 de seus atendentes, sendo medido, para cada um deles, a nota média de avaliação anterior e posterior ao treinamento e a média de gorjetas recebidas anterior e posterior ao treinamento (banco de dados “qualidade\_restaurante.csv”).

## Primeira Pergunta da Questão 3 - Letra a)

Realize análise exploratória dos dados e explicite as principais conclusões encontradas.

dadosq3 <- dadosq3[-c(1)]  
  
# Carregando o head do dataset para uma flextable  
pander::pander(head(dadosq3))

|  |  |  |  |
| --- | --- | --- | --- |
| Nota\_anterior | Nota\_posterior | Gorjeta\_anterior | Gorjeta\_posterior |
| 6.7 | 6.6 | 7,34% | 9,23% |
| 7.2 | 7.8 | 9,28% | 8,53% |
| 3.2 | 4.9 | 8,21% | 5,47% |
| 6.8 | 8.2 | 9,20% | 11,06% |
| 4.8 | 7.9 | 8,43% | 5,13% |
| 5.3 | 5.8 | 10,35% | 7,68% |

# Apresentação de informações gerais sobre as variaveis do Dataset.  
pander::pander(summary(dadosq3))

|  |  |  |  |
| --- | --- | --- | --- |
| Nota\_anterior | Nota\_posterior | Gorjeta\_anterior | Gorjeta\_posterior |
| Min. :2.9 | Min. :0.900 | Length:25 | Length:25 |
| 1st Qu.:5.1 | 1st Qu.:5.200 | Class :character | Class :character |
| Median :5.5 | Median :5.800 | Mode :character | Mode :character |
| Mean :5.8 | Mean :5.836 | NA | NA |
| 3rd Qu.:6.8 | 3rd Qu.:6.700 | NA | NA |
| Max. :8.9 | Max. :8.200 | NA | NA |

# Não há NA´s  
# Os campos de gorjeta estão como strings, vamos realizar a conversão.   
dadosq3$Gorjeta\_anterior = as.numeric(gsub("[\\%,]", "", dadosq3$Gorjeta\_anterior))/100  
dadosq3$Gorjeta\_posterior = as.numeric(gsub("[\\%,]", "", dadosq3$Gorjeta\_posterior))/100  
pander::pander(summary(dadosq3))

|  |  |  |  |
| --- | --- | --- | --- |
| Nota\_anterior | Nota\_posterior | Gorjeta\_anterior | Gorjeta\_posterior |
| Min. :2.9 | Min. :0.900 | Min. : 6.080 | Min. : 3.500 |
| 1st Qu.:5.1 | 1st Qu.:5.200 | 1st Qu.: 7.440 | 1st Qu.: 7.680 |
| Median :5.5 | Median :5.800 | Median : 8.240 | Median : 8.700 |
| Mean :5.8 | Mean :5.836 | Mean : 8.204 | Mean : 8.201 |
| 3rd Qu.:6.8 | 3rd Qu.:6.700 | 3rd Qu.: 8.760 | 3rd Qu.: 9.230 |
| Max. :8.9 | Max. :8.200 | Max. :10.350 | Max. :11.060 |

# Não há valores nulos nem valores infinitos.   
# Mais detalhes sobre a amostra  
autofit(flextable(status(dadosq3)))

| variable | q\_zeros | p\_zeros | q\_na | p\_na | q\_inf | p\_inf | type | unique |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nota\_anterior | 0 | 0 | 0 | 0 | 0 | 0 | numeric | 18 |
| Nota\_posterior | 0 | 0 | 0 | 0 | 0 | 0 | numeric | 20 |
| Gorjeta\_anterior | 0 | 0 | 0 | 0 | 0 | 0 | numeric | 24 |
| Gorjeta\_posterior | 0 | 0 | 0 | 0 | 0 | 0 | numeric | 25 |

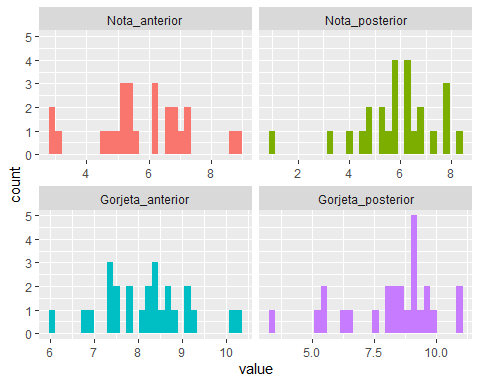
autofit(flextable(profiling\_num(dadosq3),col\_keys=c("variable","mean","std\_dev","skewness","kurtosis")))

| variable | mean | std\_dev | skewness | kurtosis |
| --- | --- | --- | --- | --- |
| Nota\_anterior | 5.8000 | 1.541104 | -0.05779144 | 2.844026 |
| Nota\_posterior | 5.8360 | 1.596319 | -1.08606631 | 4.935061 |
| Gorjeta\_anterior | 8.2044 | 1.007725 | 0.13378269 | 2.901337 |
| Gorjeta\_posterior | 8.2012 | 1.849940 | -0.85146652 | 3.141442 |

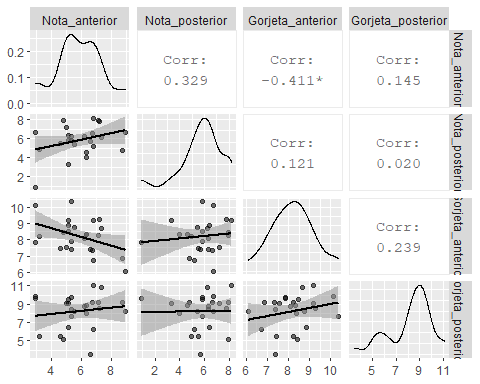
# Invertendo o DataSet para facilitar o trabalho  
  
dadosq3\_invertido <- dadosq3 %>%  
 rownames\_to\_column() %>%   
 pivot\_longer(-rowname, "key", values\_to = "value") %>%  
 mutate(atributo = ifelse(str\_detect(key,"Nota"),"Nota","Gorjeta"))  
  
mu3 <- dadosq3\_invertido %>%  
 group\_by(key) %>%  
 summarise(media = mean(value,na.rm = T))

## `summarise()` ungrouping output (override with `.groups` argument)

# Gráficos Resumidos sobre a distribuição das variáveis  
plot\_num(dadosq3, bins = 30)

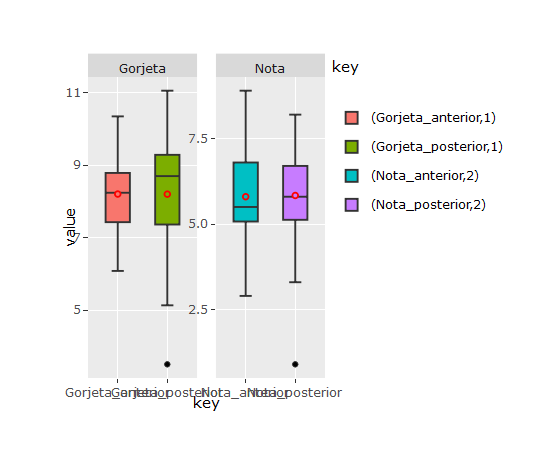


# Utilizando ggpairs para gerar informações e gráficos resumidos que auxilem da correlação entre as variáveis.  
  
ggpairs(dadosq3,columns=c("Nota\_anterior","Nota\_posterior", "Gorjeta\_anterior","Gorjeta\_posterior"),aes(alpha=0.3), lower = list(combo = wrap("facethist", bins = 30, position = "dodge"), continuous = wrap("smooth")))

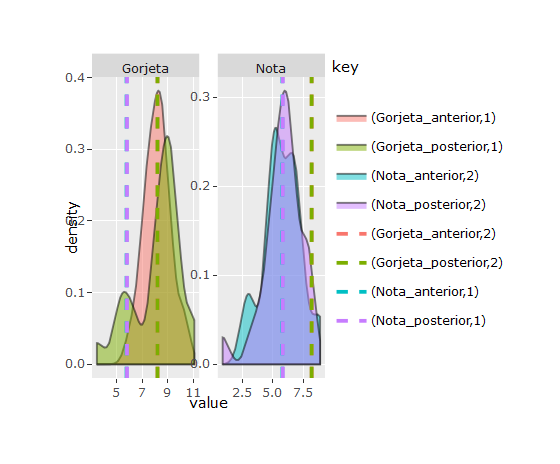


# Gráficos de caixa das Amostras  
dadosq3\_invertido %>%  
 ggplot(aes(x=key,y=value, fill = key)) %+%  
 facet\_wrap(~atributo, scales = "free") %+%  
 geom\_boxplot(show.legend = F) %+%  
 scale\_color\_brewer(palette="Dark2") %+%   
 stat\_summary(fun=mean,geom="point",color="red")%>%  
 ggplotly()

## Warning: `group\_by\_()` is deprecated as of dplyr 0.7.0.  
## Please use `group\_by()` instead.  
## See vignette('programming') for more help  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_warnings()` to see where this warning was generated.



# Gráfico de Densidade das Amostras  
dadosq3\_invertido %>%  
 ggplot(aes(x=value, fill = key)) %+%  
 geom\_density(alpha = 0.5) %+%  
 geom\_vline(data=mu3,show.legend = F, aes(xintercept = media, color = key), linetype = 2, size=1) %+%  
 facet\_wrap(~atributo, scales = "free") %>% ggplotly()



# Criando um dataset para as medidas da Nota Anterior  
#notaanterior <- data.frame("Nota\_Anterior"=dadosq3$Nota\_anterior)  
  
# Criando um dataset para as medidas da Nota Posterior  
#notaposterior <- data.frame("Nota\_Anterior"=dadosq3$Nota\_posterior)  
  
# Criando um dataset para as medidas da Gorjeta Anteior  
#gorjetaanterior <- data.frame("Nota\_Anterior"=dadosq3$Gorjeta\_anterior)  
  
# Criando um dataset para as medidas da Gorjeta Anteior  
#gorjetaposterior <- data.frame("Nota\_Anterior"=dadosq3$Gorjeta\_posterior)  
  
# Fazendo o teste para verificar a normalidade.   
  
lista <- list(nota\_anterior = dadosq3$Nota\_anterior,   
 nota\_posterior = dadosq3$Nota\_posterior,  
 gorjeta\_anterior = dadosq3$Gorjeta\_anterior,  
 gorjeta\_posterior = dadosq3$Gorjeta\_posterior)  
  
lista %>%  
 map(~shapiro.test(.x))

## $nota\_anterior  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.95633, p-value = 0.3462  
##   
##   
## $nota\_posterior  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.92198, p-value = 0.05685  
##   
##   
## $gorjeta\_anterior  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.98523, p-value = 0.9656  
##   
##   
## $gorjeta\_posterior  
##   
## Shapiro-Wilk normality test  
##   
## data: .x  
## W = 0.91794, p-value = 0.04599

lista %>%  
 map(~ks.test(.x,"pnorm",mean(.x),sd(.x)))

## Warning in ks.test(.x, "pnorm", mean(.x), sd(.x)): ties should not be present  
## for the Kolmogorov-Smirnov test

## Warning in ks.test(.x, "pnorm", mean(.x), sd(.x)): ties should not be present  
## for the Kolmogorov-Smirnov test  
  
## Warning in ks.test(.x, "pnorm", mean(.x), sd(.x)): ties should not be present  
## for the Kolmogorov-Smirnov test

## $nota\_anterior  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.1072, p-value = 0.9362  
## alternative hypothesis: two-sided  
##   
##   
## $nota\_posterior  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.11238, p-value = 0.9104  
## alternative hypothesis: two-sided  
##   
##   
## $gorjeta\_anterior  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.075509, p-value = 0.9988  
## alternative hypothesis: two-sided  
##   
##   
## $gorjeta\_posterior  
##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: .x  
## D = 0.18313, p-value = 0.3302  
## alternative hypothesis: two-sided

## Segunda Pergunta da Questão 3 - Letra b)

Há diferença estatisticamente significante entre a média de avaliação anterior e posterior ao programa de treinamentos? Considere um nível de significância de 0,05 e realize as verificações necessárias a respeito das premissas para realização deste teste de hipóteses. Explique a saída encontrada no R e interprete os resultados.

As hipóteses forma montaddas da seguinte forma:

H0: ~0 - 1=0

Ha: ~0 - 1!= 0

t.test( dadosq3$Nota\_anterior, dadosq3$Nota\_posterior, paired = T)

##   
## Paired t-test  
##   
## data: dadosq3$Nota\_anterior and dadosq3$Nota\_posterior  
## t = -0.099038, df = 24, p-value = 0.9219  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.7862188 0.7142188  
## sample estimates:  
## mean of the differences   
## -0.036

O p-value encontrado(0.9219) é maior que 0.05 não possibilitando rejeitar H0 sendo assim não é possível rejeitar a hipótese não há diferença estatística entre as médias amostrais.

## Segunda Pergunta da Questão 3 - Letra b)

Há diferença estatisticamente significante entre a gorjeta média recebida antes e após o programa de treinamentos? Considere um nível de significância de 0,05 e realize as verificações necessárias a respeito das premissas para realização deste teste de hipóteses. Explique a saída encontrada no R e interprete os resultados.

As hipóteses forma montaddas da seguinte forma

H0: 00 - 1=0

Ha: $\mi$00 - $\mi$1!= 0

t.test( dadosq3$Gorjeta\_anterior, dadosq3$Gorjeta\_posterior, paired = T)

##   
## Paired t-test  
##   
## data: dadosq3$Gorjeta\_anterior and dadosq3$Gorjeta\_posterior  
## t = 0.0084958, df = 24, p-value = 0.9933  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.7741775 0.7805775  
## sample estimates:  
## mean of the differences   
## 0.0032

O p-value(0.9933) é maior que 0.05 não permitindo rejeitar a hipótese nula significando que não é possível rejeitar a hipótese de que não há diferença estatística entre as médias amostrais.

# Questão 4

Uma consultoria deseja estimar qual a perspectiva de futuro do brasileiro para o período pós-pandemia. O objetivo é estimar a proporção de brasileiros que estão otimistas com o futuro da economia.

## Primeira Pergunta da Questão 4 - Letra a)

Considerando apenas as informações de que o nível de confiança desejado é de 95% e a margem de erro máxima esperada é de 0,03, qual o tamanho da amostra recomendado neste caso?

p <- 0.5 # Utilizando a estimativa de 0.5 pois não há informação   
E <- 0.03  
z <- qnorm(0.975)  
n <- ( (z^2) \* p \* (1 - p) ) / (E^2) # Fórmula para o cálculo do tamanho da amostra.  
n

## [1] 1067.072

ceiling(n)

## [1] 1068

O tamanho da amostra deverá ser de 1068.

## Segunda Pergunta da Questão 4 - Letra b)

Caso haja referência de uma pesquisa anterior indicando que 36% dos brasileiros estão otimistas com o futuro da economia, como os cálculos apresentados na letra a) poderiam ser revisados?

p <- 0.36 # Ao invés de utilizar os 0.5 vamos utilizar os dados da pesquisa anterior   
E <- 0.03  
z <- qnorm(0.975)  
n <- ( (z^2) \* p \* (1 - p) ) / (E^2) # Fórmula para o cálculo do tamanho da amostra.  
n

## [1] 983.4135

ceiling(n)

## [1] 984

O tamanho da amostra deverá ser de 984.