

Teste ANOVA

1, INTRODUÇÃO

1. Nesse estudo, usando o RStudio, vamos analisar um arquivo que armazena a média mensal de vendas de consultores de vendas de uma empresa fictícia
2. O arquivo VEANDAS_MEDIA_MES.csv possui dados de vendas médias mensais de 150 consultores de vendas, que pertencem à três organizações de vendas
3. Os produtos vendidos pelos vendedores nos três departamentos são exatamente os mesmos.
4. O que difere é o segmento de mercado específico para o qual cada departamento é focado
5. Os vendedores com menos de 6 meses(inclusive), são considerados novatos
6. Os demais, veteranos
7. O objetivo é determinar se existe diferença entre veteranos e novatos, no nível geral de vendas, entre departamentos e também comparar a performance dos três departamentos

2. Carregando as livrarias e pacotes

```
library(ggplot2)
library(ggpubr)
library(tidyverse)
library(broom)
library(AICcmodavg)
library(lemon)
library(knitr)
```

```
knit_print.data.frame <- lemon_print
```

3. Estabelecendo o Ambiente do programa

```
env<- "C:/data/rstudio_examples/anova"
file <- 'VENDA_MEDIA_MES.csv'
setwd(env)
getwd()
```

```
## [1] "C:/data/rstudio_examples/anova"
```

```
df<- read.csv2(file = file,header=TRUE, sep=';', dec=',' )
#View(df)
```

2.1. Criando uma coluna para dizer se o vendedor é veterano ou novato

```
df['SENIORIDADE'] <- ifelse(df['tempo_funcao']>6,'VETERANO','NOVATO' )
```

```
writeLines("td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }")
```

```
td, th { padding : 6px } th { background-color : brown ; color : white; border : 1px solid white; } td { color : brown ; border : 1px solid brown }
```

```
kable(df)
```

vendedor_id	canal_id	tempo_funcao	venda_media_mes	SENIORIDADE
V1	canal1	2	1000	NOVATO
V2	canal1	3	2000	NOVATO
V3	canal1	4	1000	NOVATO
V4	canal1	5	3000	NOVATO
V5	canal1	6	7000	NOVATO
V6	canal1	7	10000	VETERANO
V7	canal1	8	5000	VETERANO
V8	canal1	9	6000	VETERANO
V9	canal1	10	8000	VETERANO
V10	canal1	2	15000	NOVATO
V11	canal1	3	2000	NOVATO
V12	canal1	4	3000	NOVATO
V13	canal1	5	4000	NOVATO
V14	canal1	6	9000	NOVATO
V15	canal1	7	7000	VETERANO
V16	canal1	8	3000	VETERANO
V17	canal1	9	5000	VETERANO
V18	canal1	10	8000	VETERANO
V19	canal1	2	9000	NOVATO
V20	canal1	3	1000	NOVATO
V21	canal1	4	3000	NOVATO
V22	canal1	5	5000	NOVATO
V23	canal1	6	4000	NOVATO
V24	canal1	7	6000	VETERANO
V25	canal1	8	3000	VETERANO
V26	canal1	9	1000	VETERANO
V27	canal1	10	1000	VETERANO
V28	canal1	2	5000	NOVATO
V29	canal1	3	6000	NOVATO
V30	canal1	4	7000	NOVATO
V31	canal1	5	7000	NOVATO
V32	canal1	6	8000	NOVATO
V33	canal1	7	9000	VETERANO
V34	canal1	8	1000	VETERANO

vendedor_id	canal_id	tempo_funcao	venda_media_mes	SENIORIDADE
V35	canal1	9	1000	VETERANO
V36	canal1	10	10000	VETERANO
V37	canal1	2	2000	NOVATO
V38	canal1	3	3000	NOVATO
V39	canal1	4	3000	NOVATO
V40	canal1	5	3000	NOVATO
V41	canal1	6	5000	NOVATO
V42	canal1	7	7000	VETERANO
V43	canal1	8	5000	VETERANO
V44	canal1	9	5000	VETERANO
V45	canal1	10	3000	VETERANO
V46	canal1	2	1000	NOVATO
V47	canal1	3	2000	NOVATO
V48	canal1	4	1000	NOVATO
V49	canal1	5	3000	NOVATO
V50	canal1	6	7000	NOVATO
V51	canal2	7	10000	VETERANO
V52	canal2	8	5000	VETERANO
V53	canal2	9	6000	VETERANO
V54	canal2	10	8000	VETERANO
V55	canal2	2	15000	NOVATO
V56	canal2	3	2000	NOVATO
V57	canal2	4	3000	NOVATO
V58	canal2	5	4000	NOVATO
V59	canal2	6	9000	NOVATO
V60	canal2	7	7000	VETERANO
V61	canal2	8	3000	VETERANO
V62	canal2	9	5000	VETERANO
V63	canal2	10	8000	VETERANO
V64	canal2	2	9000	NOVATO
V65	canal2	3	1000	NOVATO
V66	canal2	4	3000	NOVATO
V67	canal2	5	5000	NOVATO
V68	canal2	6	4000	NOVATO
V69	canal2	7	6000	VETERANO
V70	canal2	8	3000	VETERANO
V71	canal2	9	1000	VETERANO
V72	canal2	10	1000	VETERANO
V73	canal2	2	5000	NOVATO
V74	canal2	3	6000	NOVATO
V75	canal2	4	7000	NOVATO
V76	canal2	5	7000	NOVATO
V77	canal2	6	8000	NOVATO
V78	canal2	7	9000	VETERANO
V79	canal2	8	1000	VETERANO
V80	canal2	9	1000	VETERANO
V81	canal2	10	10000	VETERANO
V82	canal2	2	2000	NOVATO
V83	canal2	3	3000	NOVATO
V84	canal2	4	3000	NOVATO
V85	canal2	5	3000	NOVATO
V86	canal2	6	5000	NOVATO

vendedor_id	canal_id	tempo_funcao	venda_media_mes	SENIORIDADE
V87	canal2	7	7000	VETERANO
V88	canal2	8	5000	VETERANO
V89	canal2	9	5000	VETERANO
V90	canal2	10	3000	VETERANO
V91	canal2	2	1000	NOVATO
V92	canal2	3	2000	NOVATO
V93	canal2	4	1000	NOVATO
V94	canal2	5	3000	NOVATO
V95	canal2	6	7000	NOVATO
V96	canal2	7	10000	VETERANO
V97	canal2	8	5000	VETERANO
V98	canal2	9	6000	VETERANO
V99	canal2	10	8000	VETERANO
V100	canal2	2	15000	NOVATO
V101	canal3	3	2000	NOVATO
V102	canal3	4	3000	NOVATO
V103	canal3	5	4000	NOVATO
V104	canal3	6	9000	NOVATO
V105	canal3	7	7000	VETERANO
V106	canal3	8	3000	VETERANO
V107	canal3	9	5000	VETERANO
V108	canal3	10	8000	VETERANO
V109	canal3	2	9000	NOVATO
V110	canal3	3	1000	NOVATO
V111	canal3	4	3000	NOVATO
V112	canal3	5	5000	NOVATO
V113	canal3	6	4000	NOVATO
V114	canal3	7	6000	VETERANO
V115	canal3	8	3000	VETERANO
V116	canal3	9	1000	VETERANO
V117	canal3	10	1000	VETERANO
V118	canal3	2	5000	NOVATO
V119	canal3	3	6000	NOVATO
V120	canal3	4	7000	NOVATO
V121	canal3	5	7000	NOVATO
V122	canal3	6	8000	NOVATO
V123	canal3	7	9000	VETERANO
V124	canal3	8	1000	VETERANO
V125	canal3	9	1000	VETERANO
V126	canal3	10	10000	VETERANO
V127	canal3	2	2000	NOVATO
V128	canal3	3	3000	NOVATO
V129	canal3	4	3000	NOVATO
V130	canal3	5	3000	NOVATO
V131	canal3	6	5000	NOVATO
V132	canal3	7	7000	VETERANO
V133	canal3	8	5000	VETERANO
V134	canal3	9	5000	VETERANO
V135	canal3	10	3000	VETERANO
V136	canal3	2	1000	NOVATO
V137	canal3	3	2000	NOVATO
V138	canal3	4	1000	NOVATO

vendedor_id	canal_id	tempo_funcao	venda_media_mes	SENIORIDADE
V139	canal3	5	3000	NOVATO
V140	canal3	6	7000	NOVATO
V141	canal3	7	10000	VETERANO
V142	canal3	8	5000	VETERANO
V143	canal3	9	6000	VETERANO
V144	canal3	10	8000	VETERANO
V145	canal3	2	15000	NOVATO
V146	canal3	3	2000	NOVATO
V147	canal3	4	3000	NOVATO
V148	canal3	5	4000	NOVATO
V149	canal3	6	9000	NOVATO
V150	canal3	7	7000	VETERANO

3. Análise preliminar dos dados

```
summary(df)
```

```
## vendedor_id      canal_id      tempo_funcao  venda_media_mes
## Length:150      Length:150      Min.   : 2.00      Min.   : 1000
## Class :character Class :character 1st Qu.: 4.00      1st Qu.: 3000
## Mode  :character Mode  :character Median : 6.00      Median : 5000
##                                     Mean  : 5.94      Mean  : 4993
##                                     3rd Qu.: 8.00      3rd Qu.: 7000
##                                     Max.   :10.00      Max.   :15000
## SENIORIDADE.tempo_funcao
## Length:150
## Class :character
## Mode  :character
##
##
##
```

3.1. Transformando variáveis em fatores

```
df$canal_id <- as.factor(df$canal_id)
df$SENIORIDADE <- as.factor(df$SENIORIDADE)
```

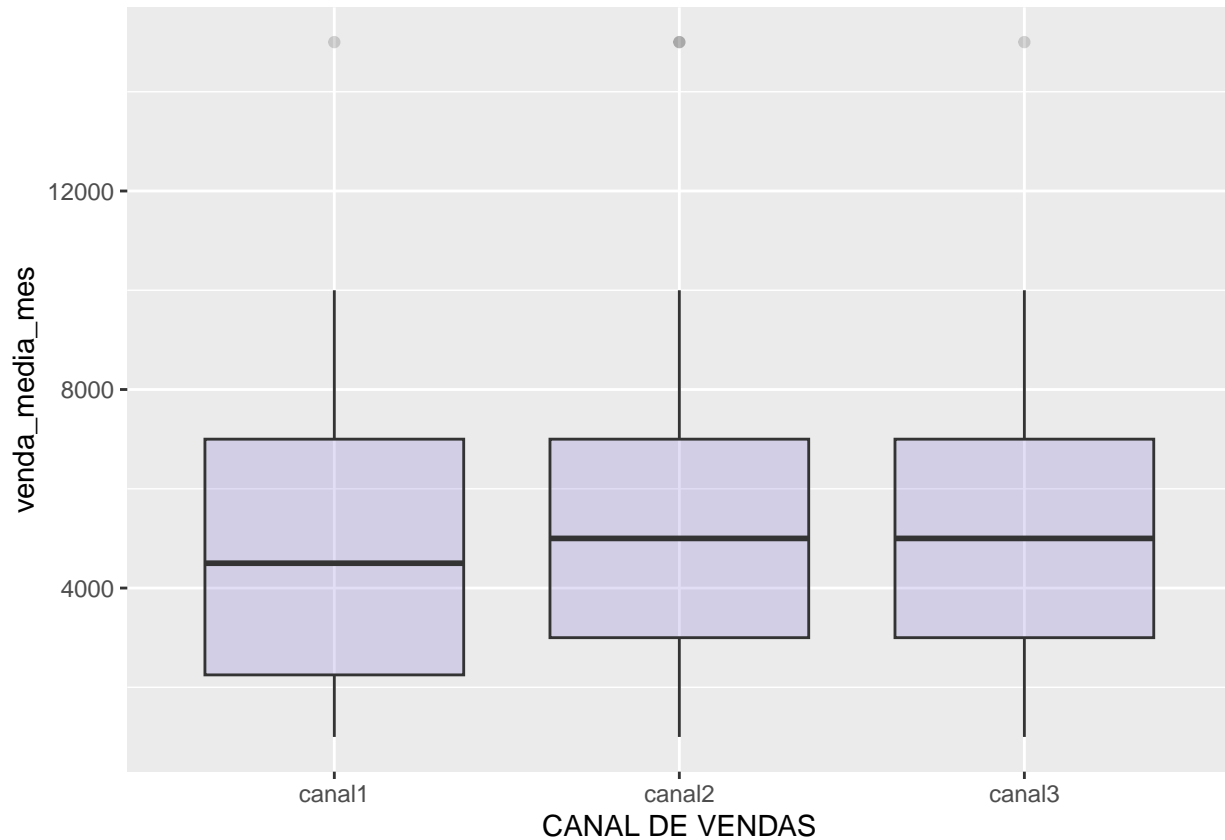
4. Análise ANOVA dos Grupos de Vendas: canal1, canal2, canal3

Para esse teste, vamos usar o ONE-WAY-ANOVA.

4.1. Desenhando o gráfico de Box Plot

Antes de iniciar a análise ANOVA, propriamente dita, vamos imprimir um gráfico de box-plot onde a variável categórica é canal_id.

```
ggplot(data=df,aes(x=as.factor(canal_id) , y=venda_media_mes ))+ geom_boxplot(fill='slateblue', alpha=0
```



O gráfico, aparentemente, mostra que não há muita diferença na média de vendas entre canais; além disso, parece haver uma variação maior nas vendas do canal 1.

4.2. ONE WAY ANOVA

```
one.way <- aov( venda_media_mes ~ canal_id, data=df )
summary(one.way)
```

```
##           Df    Sum Sq Mean Sq F value Pr(>F)
## canal_id    2  9.213e+06  4606667   0.456  0.634
## Residuals 147  1.484e+09  10093741
```

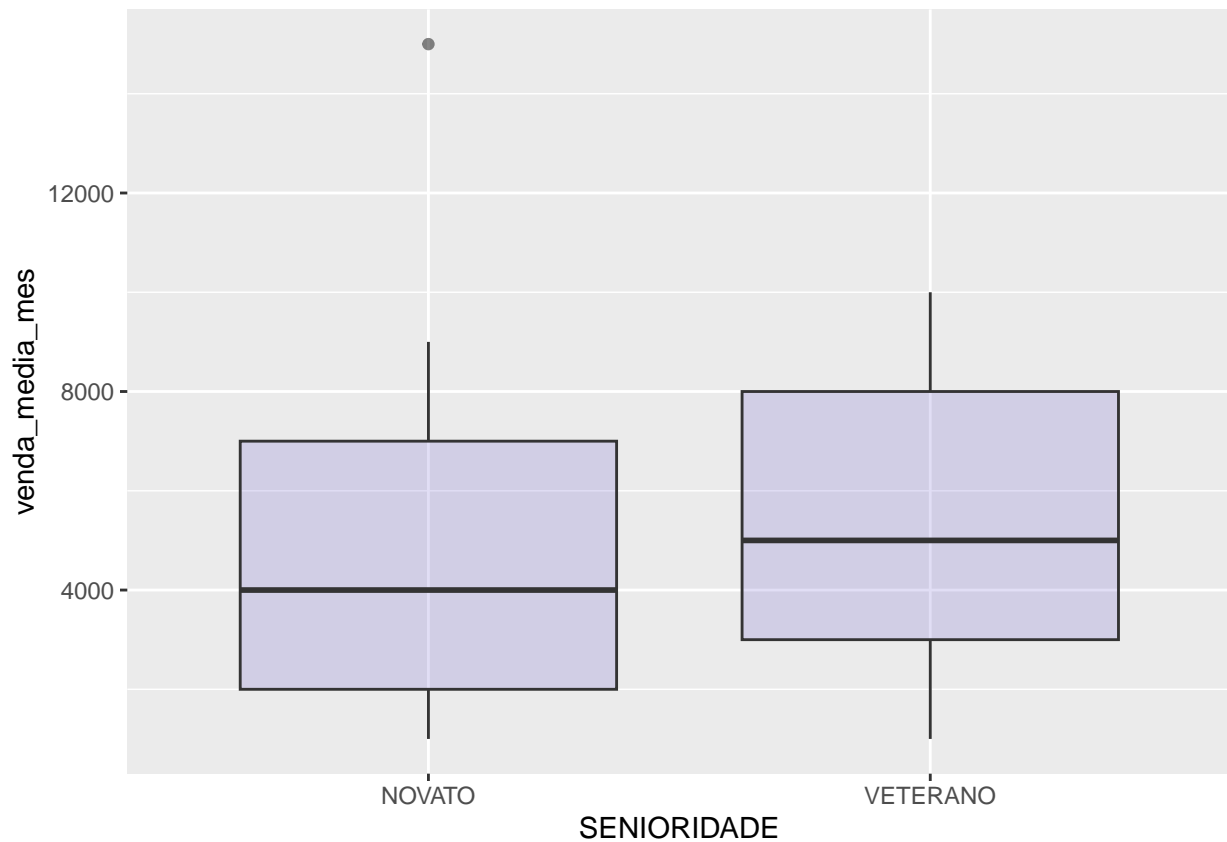
O valor de $Pr(> F_{critico})$ indica que não há diferença significativa entre a média de vendas mensais dos três canais, confirmando aquilo que já havíamos visto nos gráficos.

Vamos analisar agora se existe diferença em função da senioridade

4.3. Senioridade

Antes, vamos novamente imprimir um novo gráfico de box plot, onde a variável categórica é a senioridade

```
ggplot(data=df,aes(x=as.factor(SENIORIDADE) , y=venda_media_mes ))+ geom_boxplot(fill='slateblue', alpha=0.5)
```



Aqui, o gráfico parece indicar que existe uma diferença de desempenho entre veteranos(com uma média ligeiramente superior) e novatos.

Vamos fazer a análise anova ONEWAY

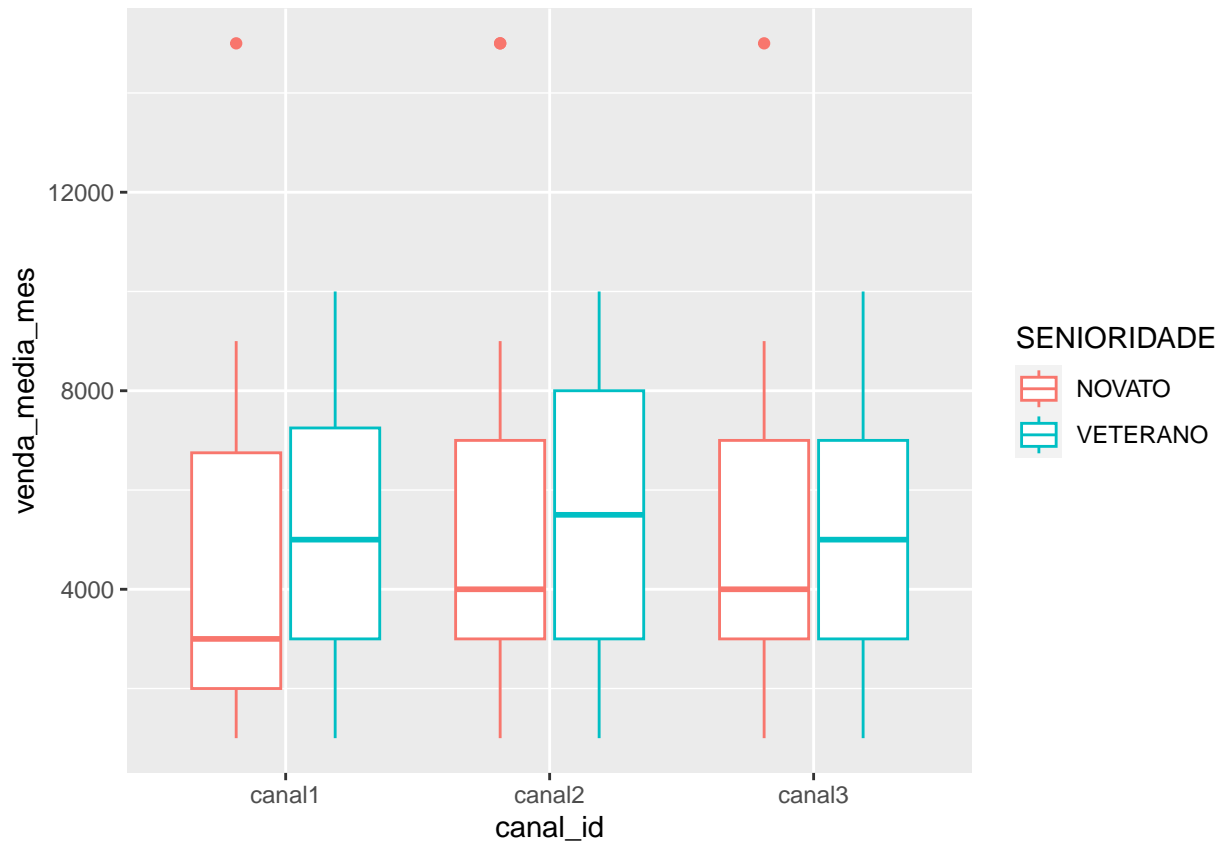
```
one.way2 <- aov( venda_media_mes ~SENIORIDADE, data=df )
summary(one.way2)
```

```
##              Df    Sum Sq Mean Sq F value Pr(>F)
## SENIORIDADE   1 1.491e+07 14908265   1.493  0.224
## Residuals  148 1.478e+09  9987061
```

O valor de p-value é de 0.224, muito acima dos 0.05 para o nível de significância de rejeição da hipótese nula. Portanto, concluímos que as médias de vendas não parecem mudar muito quando se compara veteranos e novatos.

Vamos agora, analisar os dois fatores - CANAL e Senioridade - em conjunto

```
ggplot(df, aes(x = canal_id, y =venda_media_mes , color = SENIORIDADE)) + geom_boxplot()
```



Os gráficos claramente indicam que existe diferença entre veteranos e novatos nos três canais de venda, sendo que os veteranos desempenham melhor do que os novatos em todos os canais. Além disso, conforme já verificamos anteriormente, não parece haver diferença entre os veteranos nos três canais.

Vamos realizar a análise anova two way

```
two.way <- aov( venda_media_mes ~ canal_id + SENIORIDADE, data=df)
summary(two.way)
```

```
##           Df    Sum Sq Mean Sq F value Pr(>F)
## canal_id    2 9.213e+06  4606667   0.457  0.634
## SENIORIDADE  1 1.342e+07 13419324   1.332  0.250
## Residuals 146 1.470e+09 10070964
```

5. Comparando desempenho de veteranos e novatos no canal 1

```
df_canal1 <- df %>% filter(canal_id == 'canal1')
```

```
head(df_canal1)
```

```
##   vendedor_id canal_id tempo_funcao venda_media_mes SENIORIDADE
## 1          V1  canal1           2           1000     NOVATO
## 2          V2  canal1           3           2000     NOVATO
```



```
## 3      V3   canal1      4      1000    NOVATO
## 4      V4   canal1      5      3000    NOVATO
## 5      V5   canal1      6      7000    NOVATO
## 6      V6   canal1      7     10000  VETERANO
```

```
one.way3 <- aov( venda_media_mes ~SENIORIDADE, data=df_canal1 )
summary(one.way3)
```

```
##           Df      Sum Sq Mean Sq F value Pr(>F)
## SENIORIDADE 1    7680000 7680000   0.797  0.376
## Residuals  48  462400000 9633333
```

```
df_canal2 <- df %>% filter(canal_id == 'canal2')
one.way4 <- aov( venda_media_mes ~SENIORIDADE, data=df_canal2 )
summary(one.way4)
```

```
##           Df      Sum Sq Mean Sq F value Pr(>F)
## SENIORIDADE 1    2267821  2267821   0.197  0.659
## Residuals  48  552612179 11512754
```

```
df_canal3 <- df %>% filter(canal_id == 'canal3')
one.way5 <- aov( venda_media_mes ~SENIORIDADE, data=df_canal3 )
summary(one.way5)
```

```
##           Df      Sum Sq Mean Sq F value Pr(>F)
## SENIORIDADE 1    4327389  4327389   0.457  0.502
## Residuals  48  454492611 9468596
```

Quando os canais são isolados, a análise ANOVA ONE.WAY indica que não há diferença significativa entre veteranos e novatos.