



# RELATÓRIO SOBRE TESTE T2 DE HOTELLING

JHEYMISSON THIAGO SOUSA SILVA

# Introdução

O teste  $t^2$  de Hotelling é uma extensão multivariada do teste  $t$  de Student, utilizado para comparar médias de duas amostras independentes em um contexto de múltiplas variáveis. Ele é útil quando estamos lidando com dados multivariados, nos quais as observações são medidas em várias variáveis simultaneamente. O teste  $t^2$  de Hotelling permite determinar se as médias das amostras são estatisticamente diferentes ou se podem ser consideradas iguais.

# Objetivo

O objetivo deste relatório é apresentar detalhadamente o procedimento do teste  $T^2$  de Hotelling, desde a formulação das hipóteses até a interpretação dos resultados.

# Metodologia

Formulação do teste: O teste  $t^2$  de Hotelling é baseado na estatística  $t^2$ , que segue uma distribuição F de Snedecor.

Hipóteses: As hipóteses do teste  $t^2$  de Hotelling são formuladas da seguinte maneira:

- Hipótese nula ( $H_0$ ): Não há diferenças entre as médias dos grupos.
- Hipótese alternativa ( $H_1$ ): Há diferenças implicações entre as médias dos grupos.

# Procedimento

O procedimento para realizar o teste  $t^2$  de Hotelling é o seguinte:

- ▶ Coletar os dados das variáveis dependentes nos grupos que estão sendo comparados.
- ▶ Calcular os vetores médios das variáveis dependentes em cada grupo.
- ▶ Calcular a matriz de covariância combinação das variáveis dependentes.
- ▶ Calcular uma estatística de teste  $t^2$  de Hotelling.
- ▶ Comparar o valor da estatística de teste com a distribuição F multivariada para determinar se há diferença significativa entre os grupos.

# Interpretação

Se o valor da estatística de teste  $t^2$  de Hotelling for significativo, isso indica que há uma diferença significativa entre os grupos com base nas variáveis dependentes consideradas. Nesse caso, a hipótese nula é rejeitada, e pode-se concluir que pelo menos um dos grupos tem uma média diferente das outras.

# APLICAÇÃO NO R STUDIO

BANCO DE DADOS :

<https://github.com/vincentarelbundock/Rdatasets/blob/master/csv/datasets/airquality.csv>

COM 153 OBSERVAÇÕES E 6 VARIÁVEIS .

	Ozone ↕	Solar.R ↕	Wind ↕	Temp ↕	Month ↕	Day ↕
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10
11	7	NA	6.9	74	5	11
12	16	256	9.7	69	5	12

# Importando os dados

```
library(mvtnorm)
library(ICS)
library(ICSNP) # Teste T2 Hottleing
library(MVN)

## Dados

data("airquality")
airquality

## Removendo os NA's

dados = na.omit(airquality)
dados
```

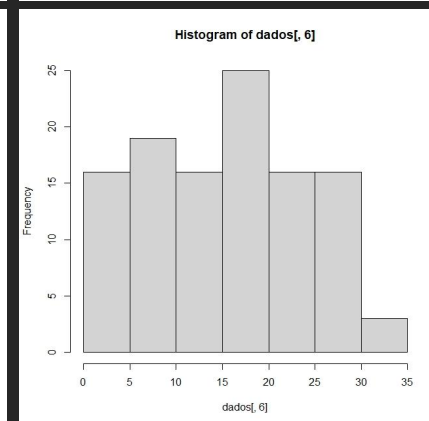
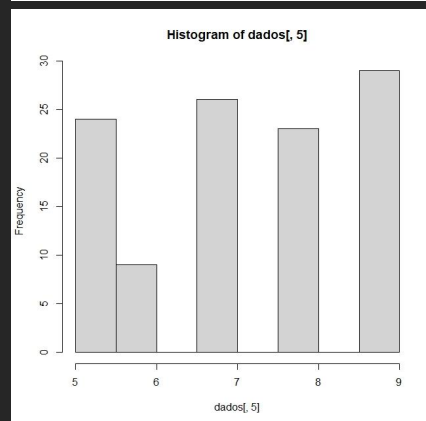
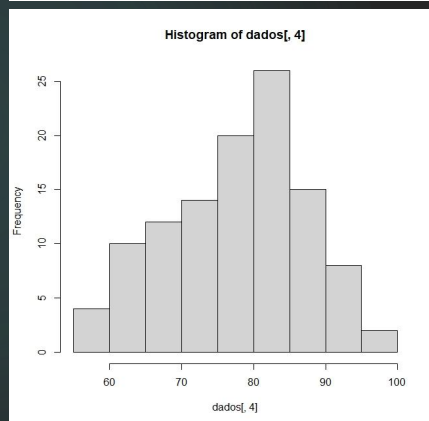
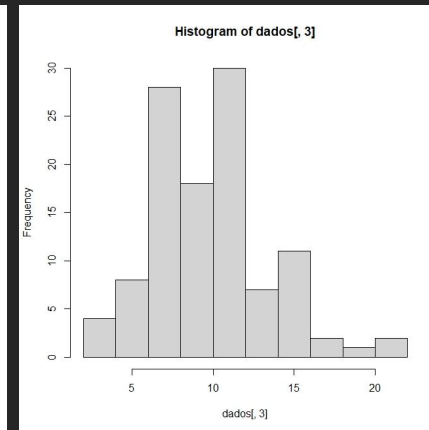
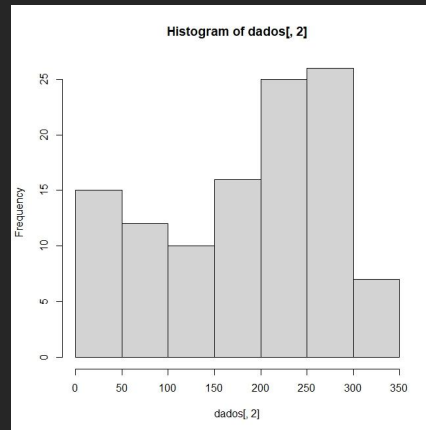
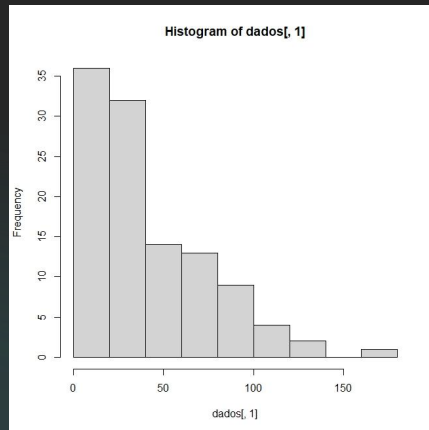


# ANÁLISE DESCRITIVA DOS DADOS

```
> summary(dados)
```

Ozone	Solar.R	Wind	Temp	Month	Day
Min. : 1.0	Min. : 7.0	Min. : 2.30	Min. : 57.00	Min. : 5.000	Min. : 1.00
1st Qu.: 18.0	1st Qu.: 113.5	1st Qu.: 7.40	1st Qu.: 71.00	1st Qu.: 6.000	1st Qu.: 9.00
Median : 31.0	Median : 207.0	Median : 9.70	Median : 79.00	Median : 7.000	Median : 16.00
Mean : 42.1	Mean : 184.8	Mean : 9.94	Mean : 77.79	Mean : 7.216	Mean : 15.95
3rd Qu.: 62.0	3rd Qu.: 255.5	3rd Qu.: 11.50	3rd Qu.: 84.50	3rd Qu.: 9.000	3rd Qu.: 22.50
Max. : 168.0	Max. : 334.0	Max. : 20.70	Max. : 97.00	Max. : 9.000	Max. : 31.00

# HISTOGRAMAS



# VERIFICANDO A SUPOSIÇÃO DE NORMALIDADE MULTIVARIADA E UNIVARIADA

```
> colMeans(dados)
      Ozone      Solar.R      wind      Temp      Month      Day
42.099099 184.801802   9.939640  77.792793   7.216216  15.945946
> ### verificar a suposição de normalidade multivariada e univariada
> mvn( dados,
+      mvnTest = 'mardia',
+      univariateTest = 'AD')
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 155.278230583413 2.87241150827957e-11 NO
2 Mardia Kurtosis 0.146098098087292 0.883843947814634 YES
3 MVN <NA> <NA> NO

$univariateNormality
      Test Variable Statistic p value Normality
1 Anderson-Darling Ozone 4.5808 <0.001 NO
2 Anderson-Darling Solar.R 2.5328 <0.001 NO
3 Anderson-Darling wind 0.6088 0.1109 YES
4 Anderson-Darling Temp 0.5988 0.1174 YES
5 Anderson-Darling Month 5.0031 <0.001 NO
6 Anderson-Darling Day 0.8502 0.0279 NO

$Descriptives
      n      Mean      Std.Dev      Median      Min      Max      25th      75th      Skew      Kurtosis
Ozone 111 42.099099 33.275969 31.0 1.0 168.0 18.0 62.0 1.23127548 1.1289943
Solar.R 111 184.801802 91.152302 207.0 7.0 334.0 113.5 255.5 -0.47969055 -0.9663681
wind 111 9.939640 3.557713 9.7 2.3 20.7 7.4 11.5 0.44949800 0.2224427
Temp 111 77.792793 9.529969 79.0 57.0 97.0 71.0 84.5 -0.22206091 -0.7098729
Month 111 7.216216 1.473434 7.0 5.0 9.0 6.0 9.0 -0.28734073 -1.2777968
Day 111 15.945946 8.707194 16.0 1.0 31.0 9.0 22.5 -0.01265914 -1.0809495
```

# CALCULANDO O TESTE T2 DE HOTELLING

```
> ## Teste T2
> ## Testar a hipotese mu = [ 42.09, 184.80, 9.93, 77.79, 7.21, 15.94 ]
> HotellingsT2(dados, Y = NULL, mu = c(42.09, 184.80, 9.93, 77.79, 7.21, 15.94), test = "f")

Hotelling's one sample T2-test

data: dados
T.2 = 0.0006818, df1 = 6, df2 = 105, p-value = 1
alternative hypothesis: true location is not equal to c(42.09,184.8,9.93,77.79,7.21,15.94)

> ## Testar a hipotese mu = [ 0, 0, 0, 0, 0, 0 ]
> HotellingsT2(dados, Y = NULL, mu = c(0,0,0,0,0,0), test = "f")

Hotelling's one sample T2-test

data: dados
T.2 = 2619.3, df1 = 6, df2 = 105, p-value < 2.2e-16
alternative hypothesis: true location is not equal to c(0,0,0,0,0,0)
```