Trabalho 1- Planejamento de Experimentos

Carlos Alberto Alves de Meneses,20180003202

2023-03-04

## Análise de Variância (ANOVA)

**Introdução**

A análise de Variância (ANOVA) é um teste estatístico considerado uma extensão do teste t amostral. A análise de variância é usada quando o número de grupos a serem comparados são dois. Se usarmos o teste t múltiplo para comparar mais de dois grupos, o teste t torna-se não confiável.

**ONE WAY ANOVA**

A análise de variância (muitas vezes chamada de ANOVA) é uma técnica estatística para analisar a maneira pela qual a média de uma variável é afetada por uma variável categórica com mais de 2 grupos.

**Hipótese**

A hipótese nula ANOVA é que todas as médias populacionais são iguais, a hipótese alternativa ANOVA é que pelo menos um par de grupos está tendo uma diferença significativa em média.

**ANOVA ONE WAY AÇÕES**

Antes de executar a ANOVA, devemos garantir que as seguintes suposições sejam atendidas:

1. Independência de observações dentro e entre amostras - Conjuntos de amostras selecionados aleatoriamente são independentes uns dos outros;
2. Normalidade da distribuição amostral - A amostra é extraída da população com distribuída normal;
3. Variância igual - Desvios padrão iguais (ou variâncias) são assumidos para as populações;
4. Quantitativo - Os dados são de natureza quantitativa.

**Objetivo**

Analisar através de exemplos, as suposições do modelo, *ONE WAY*, incluindo as comparações multiplas.

**Exemplo 01**

Para controlar a pressão sanguínea, cinco tratamentos (Placebo, T2, T3, T4, T5) incluindo um placebo, 30 pacientes (seis pacientes por grupo de tratamento) foram designados aleatoriamente. A pressão arterial de cada paciente foi medida em intervalos regulares. Há alguma melhora significativa no sangue de pacientes com diferentes tratamentos?

#One Way ANOVA para dados não empilhados #Ho: Todas as médias populacionais são iguais (μ1 = μ2 = μз.....= μn)  
#H: Pelo menos uma média é diferente das outras. (μ1 μ2 μз.....\* μn)   
  
y <- data.frame(Placebo = c(115, 118, 117, 117, 110, 108), T2=c(104,96,96,83,85,86), T3=c(86,84,84,85,84,83), T4=c(117, 124, 136, 146, 143, 145), T5=c(89, 83,85,87,87 ,86))   
y

## Placebo T2 T3 T4 T5  
## 1 115 104 86 117 89  
## 2 118 96 84 124 83  
## 3 117 96 84 136 85  
## 4 117 83 85 146 87  
## 5 110 85 84 143 87  
## 6 108 86 83 145 86

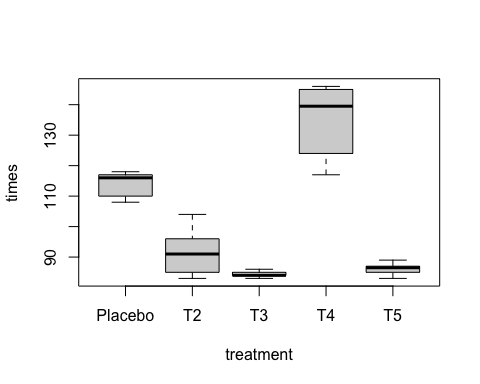
times <- stack(y)$values #empilhar os valores  
times

## [1] 115 118 117 117 110 108 104 96 96 83 85 86 86 84 84 85 84 83 117  
## [20] 124 136 146 143 145 89 83 85 87 87 86

treatment <- stack(y)$ind #pilha de variáveis indicadoras  
treatment

## [1] Placebo Placebo Placebo Placebo Placebo Placebo T2 T2 T2   
## [10] T2 T2 T2 T3 T3 T3 T3 T3 T3   
## [19] T4 T4 T4 T4 T4 T4 T5 T5 T5   
## [28] T5 T5 T5   
## Levels: Placebo T2 T3 T4 T5

boxplot(times~treatment)



**Análise de variância**

#Realize a análise de variância  
fit <- aov(times~treatment)  
fit

## Call:  
## aov(formula = times ~ treatment)  
##   
## Terms:  
## treatment Residuals  
## Sum of Squares 11503.133 1185.167  
## Deg. of Freedom 4 25  
##   
## Residual standard error: 6.88525  
## Estimated effects may be unbalanced

summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)   
## treatment 4 11503 2875.8 60.66 1.66e-12 \*\*\*  
## Residuals 25 1185 47.4   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Como o p-valor é 0,000, que é inferior a 5% do nível de significância, rejeitamos a hipótese nula, o que significa que há diferença significativa na pressão arterial média de pelo menos dois tratamentos.

res <- fit$residuals  
res

## 1 2 3 4 5 6   
## 0.8333333 3.8333333 2.8333333 2.8333333 -4.1666667 -6.1666667   
## 7 8 9 10 11 12   
## 12.3333333 4.3333333 4.3333333 -8.6666667 -6.6666667 -5.6666667   
## 13 14 15 16 17 18   
## 1.6666667 -0.3333333 -0.3333333 0.6666667 -0.3333333 -1.3333333   
## 19 20 21 22 23 24   
## -18.1666667 -11.1666667 0.8333333 10.8333333 7.8333333 9.8333333   
## 25 26 27 28 29 30   
## 2.8333333 -3.1666667 -1.1666667 0.8333333 0.8333333 -0.1666667

**Cálculo dos resíduos padronizados**

SSE = sum(res^2)  
MSE = SSE/fit$df.residual  
res.padronizado <- res/sqrt(MSE)  
round(res.padronizado, digits = 2)

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## 0.12 0.56 0.41 0.41 -0.61 -0.90 1.79 0.63 0.63 -1.26 -0.97 -0.82 0.24   
## 14 15 16 17 18 19 20 21 22 23 24 25 26   
## -0.05 -0.05 0.10 -0.05 -0.19 -2.64 -1.62 0.12 1.57 1.14 1.43 0.41 -0.46   
## 27 28 29 30   
## -0.17 0.12 0.12 -0.02

**Comparações Múltiplas**

Após realizar a análise de variância, se rejeitarmos a hipótese nula de que todas as médias amostrais são iguais, então devemos realizar uma análise Post Hoc para selecionar quais médias são significativamente diferentes das demais.

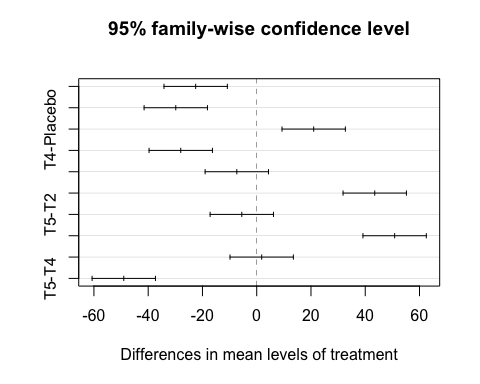
No exemplo da pressão arterial, gostaríamos de selecionar o tratamento mais eficaz, ou seja, aquele que leva à diminuição da pressão arterial.

TukeyHSD(fit)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = times ~ treatment)  
##   
## $treatment  
## diff lwr upr p adj  
## T2-Placebo -22.500000 -34.174653 -10.825347 0.0000624  
## T3-Placebo -29.833333 -41.507986 -18.158680 0.0000007  
## T4-Placebo 21.000000 9.325347 32.674653 0.0001619  
## T5-Placebo -28.000000 -39.674653 -16.325347 0.0000021  
## T3-T2 -7.333333 -19.007986 4.341320 0.3717320  
## T4-T2 43.500000 31.825347 55.174653 0.0000000  
## T5-T2 -5.500000 -17.174653 6.174653 0.6433186  
## T4-T3 50.833333 39.158680 62.507986 0.0000000  
## T5-T3 1.833333 -9.841320 13.507986 0.9901189  
## T5-T4 -49.000000 -60.674653 -37.325347 0.0000000

Os p-valores dos seguintes pares: T2-Placebo, T3-Placebo, T4-Placebo, T5 -Placebo, T4-T2, T4-T3 e T5-T4 são inferiores a 0,05, portanto, as médias são significativamente diferentes umas das outras. Portanto, a partir do box plot, podemos dizer que os tratamentos T3, T5, T2 são bastante eficazes na redução da pressão arterial em comparação com o placebo, enquanto o tratamento T4 é ineficaz na pressão arterial.

plot(TukeyHSD(fit))



*Os pares da direita são estatisticamente significativos porque não incluem o zero em seus intervalos de confiança*.

No gráfico acima, para o conjunto de comparações fornecido, os pares cujos intervalos de confiança não incluem zero são estatisticamente diferentes.

library(outliers)  
dixon.test(res)

##   
## Dixon test for outliers  
##   
## data: res  
## Q.10 = 0.33929, p-value = 0.1799  
## alternative hypothesis: lowest value -18.1666666666667 is an outlier

library(onewaytests)  
bf.test(times ~ factor(treatment), data = y)

##   
## Brown-Forsythe Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : times and factor(treatment)   
##   
## statistic : 60.662   
## num df : 4   
## denom df : 10.66322   
## p.value : 2.823521e-07   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

out <- bf.test(times ~ factor(treatment), data = y)

##   
## Brown-Forsythe Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : times and factor(treatment)   
##   
## statistic : 60.662   
## num df : 4   
## denom df : 10.66322   
## p.value : 2.823521e-07   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

#paircomp(out)

A ANOVA também é considerada um modelo linear e, portanto, também podemos realizar a análise abaixo:

model <- lm(times~treatment)  
model

##   
## Call:  
## lm(formula = times ~ treatment)  
##   
## Coefficients:  
## (Intercept) treatmentT2 treatmentT3 treatmentT4 treatmentT5   
## 114.17 -22.50 -29.83 21.00 -28.00

anova(model)

## Analysis of Variance Table  
##   
## Response: times  
## Df Sum Sq Mean Sq F value Pr(>F)   
## treatment 4 11503.1 2875.78 60.662 1.662e-12 \*\*\*  
## Residuals 25 1185.2 47.41   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(model)

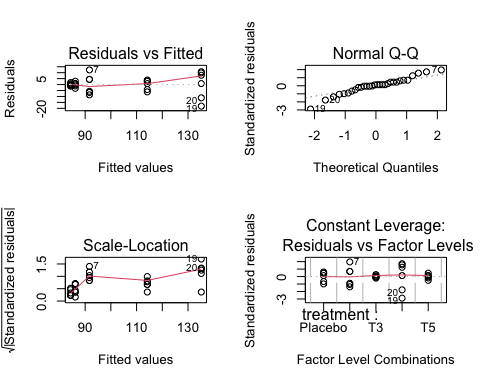
##   
## Call:  
## lm(formula = times ~ treatment)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.167 -2.708 0.750 2.833 12.333   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 114.167 2.811 40.616 < 2e-16 \*\*\*  
## treatmentT2 -22.500 3.975 -5.660 6.83e-06 \*\*\*  
## treatmentT3 -29.833 3.975 -7.505 7.38e-08 \*\*\*  
## treatmentT4 21.000 3.975 5.283 1.80e-05 \*\*\*  
## treatmentT5 -28.000 3.975 -7.044 2.21e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.885 on 25 degrees of freedom  
## Multiple R-squared: 0.9066, Adjusted R-squared: 0.8916   
## F-statistic: 60.66 on 4 and 25 DF, p-value: 1.662e-12

**Validação das suposições**

**Normalidade dos resíduos**

Como o modelo ANOVA corresponde a um modelo linear com uma variável independente categórica. Assim, os pressupostos das regressões podem ser validados pelo método de análise de resíduos.

# mfrow significa simplesmente "layout de linha de vários quadros". mfcol significa WitiFrame layout por coluna".   
par(mfrow = c(2,2))   
plot(fit)



* Teste Shapiro Wilk - Normalidade dos resíduos

shapiro.test(res)

##   
## Shapiro-Wilk normality test  
##   
## data: res  
## W = 0.95545, p-value = 0.236

**Homogeneidade da variância [Homoscedasticidade]**

* Teste de Homogeneidade de variância

library(stats)  
fligner.test(times~treatment, data = y)

##   
## Fligner-Killeen test of homogeneity of variances  
##   
## data: times by treatment  
## Fligner-Killeen:med chi-squared = 16.483, df = 4, p-value = 0.002435

Assume-se que as populações têm desvios padrão (ou variâncias) iguais. Na ANOVA também é possível estimar se a hipótese de homocedasticidade é admissível.

#Ho: Homocedasticidade   
#Hi: Heteroscedasticidade   
bartlett.test(times~treatment)

##   
## Bartlett test of homogeneity of variances  
##   
## data: times by treatment  
## Bartlett's K-squared = 26.571, df = 4, p-value = 2.427e-05

Como o p-valor é 0,000, que é menor que 5% do nível de significância, rejeitamos a hipótese nula e concluímos que as variâncias de cinco tratamentos não são iguais. Mas o teste de Bartlett não é tão robusto para população não normal e, portanto, em caso de não normalidade usamos o Teste de Levene.

#Ho: Homoscedasticity  
#H: Heteroscedasticity  
#install.packages("car")   
library(car) #Para o Teste de Levene você precisa do pacote "car"

## Loading required package: carData

leveneTest(times, treatment)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 4 5.1778 0.003529 \*\*  
## 25   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

* Teste Robusto para a anova oneway

library(onewaytests)  
welch.test(times~treatment, data = y)

##   
## Welch's Heteroscedastic F Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : times and treatment   
##   
## statistic : 82.75572   
## num df : 4   
## denom df : 11.1254   
## p.value : 3.374512e-08   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

#muito Robusto contra desvios da normalidade   
bartlett.test(times ~ treatment, data=y)

##   
## Bartlett test of homogeneity of variances  
##   
## data: times by treatment  
## Bartlett's K-squared = 26.571, df = 4, p-value = 2.427e-05

# Teste de Fligner-Killeen Teste não paramétrico. Muito robusto contra desvios da normalidade   
fligner.test(times ~ treatment, data=y)

##   
## Fligner-Killeen test of homogeneity of variances  
##   
## data: times by treatment  
## Fligner-Killeen:med chi-squared = 16.483, df = 4, p-value = 0.002435

Como o p-valor é 0,00243, que é menor que o nível de significância de 5%, rejeitamos a hipótese nula e concluímos que as variâncias de cinco tratamentos não são iguais.

**Exemplo 02**

Um economista queria comparar o consumo de eletricidade em quatro grandes cidades da Austrália. Ele pegou amostras aleatórias de 5 domicílios de duas pessoas de cada cidade (1-Adelaide, 2=Hobart, 3-Perth, 4-Melbourne) e seu consumo de energia.

**Solução**

Utilizaremos o banco *electricity* como nosso banco de dados e importaremos esse arquivo para o R.

#ANOVA de uma via para dados empilhados #Ho: Não há diferença significativa no Principais cidades da Austrália Consumo de energia de quatro diferenças em pelo menos duas das quatro principais cidades da significativa Austrália  
library(readr)  
electricity <- read\_delim("~/Documents/ESTATISTICA/PLANEJAMENTO/electricity.txt",delim = "\t", escape\_double = FALSE,trim\_ws = TRUE)

## Rows: 20 Columns: 2  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: "\t"  
## dbl (2): city, consumption  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

attach(electricity)   
city1 = as.factor(city)   
summary(city1)

## 1 2 3 4   
## 5 5 5 5

#convertendo valor numérico em valor categórico  
city1 = factor(city1, labels=c("Adelaide", "Hobart", "Melbourne", "Perth"))  
summary(city1)

## Adelaide Hobart Melbourne Perth   
## 5 5 5 5

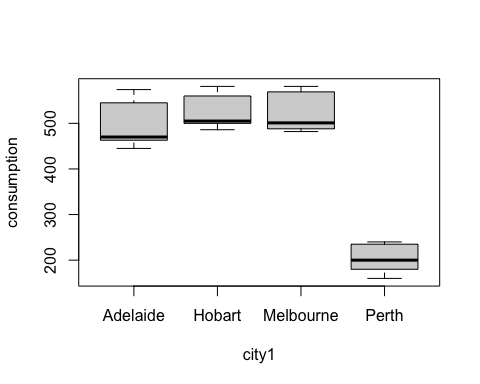
fit = aov (consumption~city1)   
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)   
## city1 3 371199 123733 59.39 6.83e-09 \*\*\*  
## Residuals 16 33333 2083   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Desde que o p- valor é 0,000, que é inferior a 5% do nível de significância, rejeitamos a hipótese nula, o que significa que há uma diferença significativa no consumo de eletricidade das quatro principais cidades da Austrália.

**Checagem do Modelo**

boxplot(consumption~city1)



res <- fit$residuals  
res

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## 45.6 -29.4 -54.4 74.6 -36.4 54.6 -26.4 -40.4 33.6 -21.4 56.8 -42.2 -23.2   
## 14 15 16 17 18 19 20   
## 44.8 -36.2 37.0 32.0 -3.0 -43.0 -23.0

* Cálculo dos resíduos padronizados

SSE = sum(res^2)  
MSE = SSE/fit$df.residual  
resid.padronizado <- res/sqrt(MSE)  
round(resid.padronizado, digits = 2)

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## 1.00 -0.64 -1.19 1.63 -0.80 1.20 -0.58 -0.89 0.74 -0.47 1.24 -0.92 -0.51   
## 14 15 16 17 18 19 20   
## 0.98 -0.79 0.81 0.70 -0.07 -0.94 -0.50

library(outliers)  
dixon.test(res)

##   
## Dixon test for outliers  
##   
## data: res  
## Q.4 = 0.17123, p-value = 0.6674  
## alternative hypothesis: highest value 74.6 is an outlier

Teste da Normalidade dos Resíduos

shapiro.test(res)

##   
## Shapiro-Wilk normality test  
##   
## data: res  
## W = 0.87459, p-value = 0.01416

* Teste de Homogeneidade de variância

library(lawstat)

##   
## Attaching package: 'lawstat'

## The following object is masked from 'package:car':  
##   
## levene.test

leveneTest(consumption, city1)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 3 0.1828 0.9065  
## 16

bartlett.test(consumption,city1)

##   
## Bartlett test of homogeneity of variances  
##   
## data: consumption and city1  
## Bartlett's K-squared = 0.92271, df = 3, p-value = 0.8199

library(stats)  
fligner.test(consumption~city1, data = electricity)

##   
## Fligner-Killeen test of homogeneity of variances  
##   
## data: consumption by city1  
## Fligner-Killeen:med chi-squared = 0.51912, df = 3, p-value = 0.9147

**Teste robusto para a anova oneway**

library(onewaytests)  
welch.test(consumption~city1, data = electricity)

##   
## Welch's Heteroscedastic F Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : consumption and city1   
##   
## statistic : 74.91338   
## num df : 3   
## denom df : 8.766704   
## p.value : 1.421872e-06   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

* Teste de Bryan Forsythe

bf.test(consumption~factor(city1), data = electricity)

##   
## Brown-Forsythe Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : consumption and factor(city1)   
##   
## statistic : 59.39201   
## num df : 3   
## denom df : 14.21597   
## p.value : 2.777317e-08   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

out6 <- bf.test(consumption~city1, data = electricity)

##   
## Brown-Forsythe Test (alpha = 0.05)   
## -------------------------------------------------------------   
## data : consumption and city1   
##   
## statistic : 59.39201   
## num df : 3   
## denom df : 14.21597   
## p.value : 2.777317e-08   
##   
## Result : Difference is statistically significant.   
## -------------------------------------------------------------

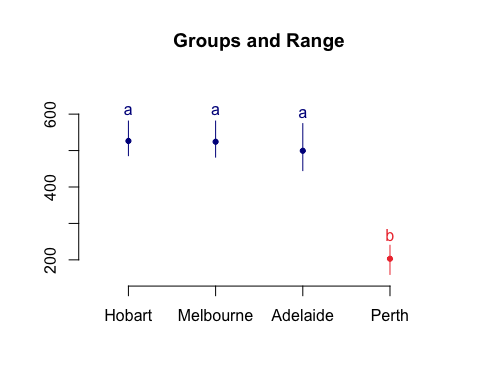
paircomp(out6)

##   
## Bonferroni Correction (alpha = 0.05)   
## -----------------------------------------------------   
## Level (a) Level (b) p.value No difference  
## 1 Adelaide Hobart 1.000000e+00 Not reject  
## 2 Adelaide Melbourne 1.000000e+00 Not reject  
## 3 Adelaide Perth 1.852681e-04 Reject  
## 4 Hobart Melbourne 1.000000e+00 Not reject  
## 5 Hobart Perth 7.496003e-06 Reject  
## 6 Melbourne Perth 2.195484e-05 Reject  
## -----------------------------------------------------

library(onewaytests)  
library(agricolae)  
out7 <- HSD.test(fit, "city1")  
out7

## $statistics  
## MSerror Df Mean CV MSD  
## 2083.325 16 438.25 10.41494 82.59036  
##   
## $parameters  
## test name.t ntr StudentizedRange alpha  
## Tukey city1 4 4.046093 0.05  
##   
## $means  
## consumption std r Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 445 574 463 470 545  
## Hobart 526.4 41.52469 5 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 482 581 488 501 569  
## Perth 203.0 34.56877 5 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

plot(out7)



out8 <- LSD.test(fit, "city1")  
out8

## $statistics  
## MSerror Df Mean CV t.value LSD  
## 2083.325 16 438.25 10.41494 2.119905 61.19627  
##   
## $parameters  
## test p.ajusted name.t ntr alpha  
## Fisher-LSD none city1 4 0.05  
##   
## $means  
## consumption std r LCL UCL Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 456.1277 542.6723 445 574 463 470 545  
## Hobart 526.4 41.52469 5 483.1277 569.6723 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 480.9277 567.4723 482 581 488 501 569  
## Perth 203.0 34.56877 5 159.7277 246.2723 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

out9 <- scheffe.test(fit, "city1")  
out9

## $statistics  
## MSerror Df F Mean CV Scheffe CriticalDifference  
## 2083.325 16 3.238872 438.25 10.41494 3.117148 89.98415  
##   
## $parameters  
## test name.t ntr alpha  
## Scheffe city1 4 0.05  
##   
## $means  
## consumption std r Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 445 574 463 470 545  
## Hobart 526.4 41.52469 5 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 482 581 488 501 569  
## Perth 203.0 34.56877 5 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

comparison <- duncan.test(fit, "city1")  
comparison

## $statistics  
## MSerror Df Mean CV  
## 2083.325 16 438.25 10.41494  
##   
## $parameters  
## test name.t ntr alpha  
## Duncan city1 4 0.05  
##   
## $duncan  
## Table CriticalRange  
## 2 2.997999 61.19627  
## 3 3.143802 64.17247  
## 4 3.234945 66.03290  
##   
## $means  
## consumption std r Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 445 574 463 470 545  
## Hobart 526.4 41.52469 5 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 482 581 488 501 569  
## Perth 203.0 34.56877 5 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

out10 <- SNK.test(fit, "city1")  
out10

## $statistics  
## MSerror Df Mean CV  
## 2083.325 16 438.25 10.41494  
##   
## $parameters  
## test name.t ntr alpha  
## SNK city1 4 0.05  
##   
## $snk  
## Table CriticalRange  
## 2 2.997999 61.19627  
## 3 3.649139 74.48759  
## 4 4.046093 82.59036  
##   
## $means  
## consumption std r Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 445 574 463 470 545  
## Hobart 526.4 41.52469 5 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 482 581 488 501 569  
## Perth 203.0 34.56877 5 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

out11 <- waller.test(fit, "city1")  
out11

## $statistics  
## Mean Df CV MSerror F.Value Waller CriticalDifference  
## 438.25 16 10.41494 2083.325 59.39201 1.922 55.48325  
##   
## $parameters  
## test name.t ntr K  
## Waller-Duncan city1 4 100  
##   
## $means  
## consumption std r Min Max Q25 Q50 Q75  
## Adelaide 499.4 56.55351 5 445 574 463 470 545  
## Hobart 526.4 41.52469 5 486 581 500 505 560  
## Melbourne 524.2 47.07122 5 482 581 488 501 569  
## Perth 203.0 34.56877 5 160 240 180 200 235  
##   
## $comparison  
## NULL  
##   
## $groups  
## consumption groups  
## Hobart 526.4 a  
## Melbourne 524.2 a  
## Adelaide 499.4 a  
## Perth 203.0 b  
##   
## attr(,"class")  
## [1] "group"

library(DescTools)

##   
## Attaching package: 'DescTools'

## The following object is masked from 'package:car':  
##   
## Recode

DunnettTest(consumption, city1, control = NULL, conf.level = 0.95)

##   
## Dunnett's test for comparing several treatments with a control :   
## 95% family-wise confidence level  
##   
## $Adelaide  
## diff lwr.ci upr.ci pval   
## Hobart-Adelaide 27.0 -47.84207 101.84207 0.6775   
## Melbourne-Adelaide 24.8 -50.04207 99.64207 0.7275   
## Perth-Adelaide -296.4 -371.24207 -221.55793 2.6e-10 \*\*\*  
##   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Cálculo das médias uando a função tapply**

#usando a função tapply para calcular a média de quatro cidades   
tapply(consumption, city1, mean)

## Adelaide Hobart Melbourne Perth   
## 499.4 526.4 524.2 203.0

Pode-se observar que o consumo de eletricidade em Perth é muito menor do que nas outras três cidades da Austrália. Isso pode ser confirmado usando o teste PostHoc Tukey.

TukeyHSD (fit)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = consumption ~ city1)  
##   
## $city1  
## diff lwr upr p adj  
## Hobart-Adelaide 27.0 -55.59036 109.59036 0.7866225  
## Melbourne-Adelaide 24.8 -57.79036 107.39036 0.8254692  
## Perth-Adelaide -296.4 -378.99036 -213.80964 0.0000001  
## Melbourne-Hobart -2.2 -84.79036 80.39036 0.9998341  
## Perth-Hobart -323.4 -405.99036 -240.80964 0.0000000  
## Perth-Melbourne -321.2 -403.79036 -238.60964 0.0000000

O p-valor dos seguintes pares: Perth-Adelaide, Perth-Hobart, Perth-Melbourne são inferiores a 4,05, podemos dizer que há uma diferença significativa no consumo de eletricidade nesses pares.

plot(TukeyHSD (fit))



No gráfico acima, para o conjunto de comparações fornecido, os pares cujo intervalo de confiança não incluem o zero são estatisticamente diferentes.

**ONE WAY ANOVA USANDO O PACOTE [“userfriendlyscience”]**

Como um acrescimo no nosso conhecimento, apresento a seguir um pacote que nos auxilia nas nossas análises.

Infelizmente, esse pacote não está mais disponível no CRAN-R e foi baixado através da página dos próprios desenvolvedores.

O crédito pelo desenvolvimento de um pacote amigável à ciência vai para Gjalt-Jorn Peters, Peter Verboon, James Green. O pacote também abriga convenientemente uma série de funções adicionais destinadas a aumentar a qualidade da metodologia e estatística em psicologia, não oferecendo soluções técnicas, mas mudando as perspectivas, por exemplo, em direção ao raciocínio baseado em distribuições de amostragem em oposição a estimativas pontuais.

#Instalando dependências  
#devtools::install\_github("matherion/userfriendlyscience", dependencies=TRUE)  
#install.packages("userfriendlyscience")  
library(userfriendlyscience)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

oneway(consumption, city1, posthoc= "games-howell")

## ### Oneway Anova for y=consumption and x=city1 (groups: Adelaide, Hobart, Melbourne, Perth)

## Registered S3 methods overwritten by 'ufs':  
## method from   
## grid.draw.ggProportionPlot userfriendlyscience  
## pander.associationMatrix userfriendlyscience  
## pander.dataShape userfriendlyscience  
## pander.descr userfriendlyscience  
## pander.normalityAssessment userfriendlyscience  
## print.CramersV userfriendlyscience  
## print.associationMatrix userfriendlyscience  
## print.confIntOmegaSq userfriendlyscience  
## print.confIntV userfriendlyscience  
## print.dataShape userfriendlyscience  
## print.descr userfriendlyscience  
## print.ggProportionPlot userfriendlyscience  
## print.meanConfInt userfriendlyscience  
## print.multiVarFreq userfriendlyscience  
## print.normalityAssessment userfriendlyscience  
## print.regrInfluential userfriendlyscience  
## print.scaleDiagnosis userfriendlyscience  
## print.scaleStructure userfriendlyscience  
## print.scatterMatrix userfriendlyscience

## Omega squared: 95% CI = [.77; .94], point estimate = .9  
## Eta Squared: 95% CI = [.8; .94], point estimate = .92  
##   
## SS Df MS F p  
## Between groups (error + effect) 371198.55 3 123732.85 59.39 <.001  
## Within groups (error only) 33333.2 16 2083.32   
##   
##   
## ### Post hoc test: games-howell  
##   
## diff ci.lo ci.hi t df p  
## Hobart-Adelaide 27.0 -75.59 129.59 0.86 7.34 .825  
## Melbourne-Adelaide 24.8 -81.38 130.98 0.75 7.74 .873  
## Perth-Adelaide -296.4 -396.02 -196.78 10.00 6.62 <.001  
## Melbourne-Hobart -2.2 -92.42 88.02 0.08 7.88 1.000  
## Perth-Hobart -323.4 -401.37 -245.43 13.38 7.75 <.001  
## Perth-Melbourne -321.2 -406.59 -235.81 12.30 7.34 <.001

#2 Anova unidirecional para y=consumo e x-cidade1 (grupos: Adelaide, Hobart, Melbourne, Perth )

Como o p-valor é 0,001, que é inferior a 5% do nível de significância, rejeitamos a hipótese nula, que significa que há uma diferença significativa no consumo de eletricidade das quatro maiores cidades da Austrália. Além disso, o valor p dos seguintes pares Perth-Adelaide, Perth-Hobart, Perth-Melbourne é menor que 0,05, então podemos dizer que há uma diferença significativa no consumo de eletricidade neste par. O valor p de Hobart-Adelaide e Melbourne-Adelaide é 0,825 e 0,873, que é maior que 0,05, portanto não rejeitamos a hipótese nula e concluímos que não há diferença significativa no consumo de eletricidade nesses pares.