> data1<-read.csv("energy.csv", header=T)

> data1

consumption area

1 13 Northeast

2 8 Northeast

3 11 Northeast

4 12 Northeast

5 11 Northeast

6 15 Midwest

7 10 Midwest

8 16 Midwest

9 11 Midwest

10 13 Midwest

11 10 Midwest

12 5 South

13 11 South

14 9 South

15 5 South

16 7 South

17 8 West

18 10 West

19 6 West

20 5 West

>

> n = length(data1$area)

>

> # var1 - nr of observations

> ng = table(data1$area)

> ng

Midwest Northeast South West

6 5 5 4

>

> # var2 - nr of group observations

> ng2 = tapply(data1$consumption, data1$area, length)

> ng2

Midwest Northeast South West

6 5 5 4

>

> # media globala

> mg = mean(data1$consumption)

>

> # var1 - group mean

> mgg = tapply(data1$consumption, data1$area, mean)

> mgg

Midwest Northeast South West

12.50 11.00 7.40 7.25

>

> # var2 - group mean

> mgg2 = aggregate(data1$consumption, by = list(data1$area), mean)

> mgg2

Group.1 x

1 Midwest 12.50

2 Northeast 11.00

3 South 7.40

4 West 7.25

>

> # varianta 1 - group Standard deviation

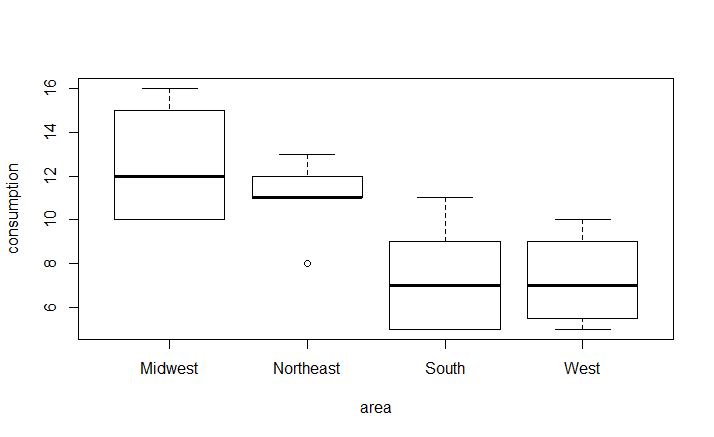
> sgg = tapply(data1$consumption, data1$area, sd)

> sgg

Midwest Northeast South West

2.588436 1.870829 2.607681 2.217356

plot(consumption~area, data = data1)



> #One way ANOVA model

> anova\_model = aov(consumption~area, data = data1)

>

> summary(anova\_model)

Df Sum Sq Mean Sq F value Pr(>F)

area 3 105.75 35.25 6.305 0.00499 \*\*

Residuals 16 89.45 5.59

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> shapiro.test(residuals(anova\_model))

Shapiro-Wilk normality test

data: residuals(anova\_model)

W = 0.92745, p-value = 0.1379

##

> bartlett.test(consumption~area, data = data1)

Bartlett test of homogeneity of variances

data: consumption by area

Bartlett's K-squared = 0.51476, df = 3, p-value = 0.9156

> library(car)

> leveneTest(y=data1$consumption, group=data1$area)

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 3 0.6161 0.6145

16

oneway.test(consumption~area, data = data1, var.equal = T)

One-way analysis of means

data: consumption and area

F = 6.3052, num df = 3, denom df = 16, p-value = 0.004994

##

> model = lm(consumption~area, data = data1)

> anova(model)

Analysis of Variance Table

Response: consumption

Df Sum Sq Mean Sq F value Pr(>F)

area 3 105.75 35.250 6.3052 0.004994 \*\*

Residuals 16 89.45 5.591

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**SCRIPT**

data1<-read.csv("energy.csv", header=T)

data1

n = length(data1$area)

# var1 - nr of observations

ng = table(data1$area)

ng

# var2 - nr of group observations

ng2 = tapply(data1$consumption, data1$area, length)

ng2

# media globala

mg = mean(data1$consumption)

# var1 - group mean

mgg = tapply(data1$consumption, data1$area, mean)

mgg

# var2 - group mean

mgg2 = aggregate(data1$consumption, by = list(data1$area), mean)

mgg2

# varianta 1 - group Standard deviation

sgg = tapply(data1$consumption, data1$area, sd)

sgg

plot(consumption~area, data = data1)

######

#One way ANOVA model

anova\_model = aov(consumption~area, data = data1)

summary(anova\_model)

#Shapiro–Wilk test tests the null hypothesis that a sample x1, ..., xn came from a normally distributed population

shapiro.test(residuals(anova\_model))

#Bartlett test to test the null hypothesis of equal group variances.

bartlett.test(consumption~area, data = data1)

######

library(car)

leveneTest(y=data1$consumption, group=data1$area)

oneway.test(consumption~area, data = data1, var.equal = T)

######

model = lm(consumption~area, data = data1)

anova(model)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| >##Example2  > setwd("C:/Mihaela/R Scripts")  > data2<-read.csv("bank.csv", header=T)  > data2  no\_customers teller  1 19 A  2 21 A  3 26 A  4 24 A  5 18 A  6 14 B  7 16 B  8 14 B  9 13 B  10 17 B  11 13 B  12 11 C  13 14 C  14 21 C  15 13 C  16 16 C  17 18 C  18 24 D  19 19 D  20 21 D  21 26 D  22 20 D  >  > n = length(data2$teller)  > ng = table(data2$teller)  > ng  A B C D  5 6 6 5  >  > plot(no\_customers~teller, data = data2)  >  > qqnorm(data2$no\_customers)  > qqline(data2$no\_customers)  >    > par(mfrow=c(1,2)) # set graphics window to plot side-by-side  > plot(anova\_model, 1) # graphical test of homogeneity  > plot(anova\_model, 2) # graphical test of normality  > par(mfrow=c(1,1))  >    > #One way ANOVA model  > anova\_model = aov(no\_customers~teller, data = data2)  >  > #Shapiro–Wilk test tests the null hypothesis that a sample x1, ..., xn came from a normally distributed population  > shapiro.test(residuals(anova\_model)) #test for normality  Shapiro-Wilk normality test  data: residuals(anova\_model)  W = 0.96009, p-value = 0.4911  >  > library(car)  > leveneTest(y=data2$no\_customers, group=data2$teller) #test for homogeneity of variances  Levene's Test for Homogeneity of Variance (center = median)  Df F value Pr(>F)  group 3 1.074 0.385  18  >  >  > summary(anova\_model) #ANOVA table  Df Sum Sq Mean Sq F value Pr(>F)  teller 3 255.6 85.21 9.695 0.000498 \*\*\*  Residuals 18 158.2 8.79  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  >  >  > boxplot(data2$no\_customers~data2$teller, ylab = c("Number of Customers"), xlab=c("Teller"))    > library(asbio)  > library(tcltk)  >  > pairw.anova(data2$no\_customers,data2$teller, method = "tukey", conf.level = 0.95)  95% Tukey-Kramer confidence intervals  Diff Lower Upper Decision Adj. p-value  muA-muB 7.1 2.02636 12.17364 Reject H0 0.004683  muA-muC 6.1 1.02636 11.17364 Reject H0 0.01543  muB-muC -1 -5.83752 3.83752 FTR H0 0.935546  muA-muD -0.4 -5.69924 4.89924 FTR H0 0.996441  muB-muD -7.5 -12.57364 -2.42636 Reject H0 0.002893  muC-muD -6.5 -11.57364 -1.42636 Reject H0 0.009609  >  > pairw.anova(data2$no\_customers,data2$teller,method = "bonf", conf.level = 0.95)  95% Bonferroni confidence intervals  Diff Lower Upper Decision Adj. p-value  muA-muB 7.1 1.78142 12.41858 Reject H0 0.005569  muA-muC 6.1 0.78142 11.41858 Reject H0 0.019233  muB-muC -1 -6.07106 4.07106 FTR H0 1  muA-muD -0.4 -5.95507 5.15507 FTR H0 1  muB-muD -7.5 -12.81858 -2.18142 Reject H0 0.003392  muC-muD -6.5 -11.81858 -1.18142 Reject H0 0.011726  >  > pairw.anova(data2$no\_customers,data2$teller,method = "scheffe", conf.level = 0.95)  95% Scheffe confidence intervals  Diff Lower Upper Decision Adj. P-value  muA-muB 7.1 1.57286 12.62714 Reject H0 0.009107  muA-muC 6.1 0.57286 11.62714 Reject H0 0.027327  muB-muC -1 -6.26992 4.26992 FTR H0 0.950898  muA-muD -0.4 -6.17291 5.37291 FTR H0 0.997354  muB-muD -7.5 -13.02714 -1.97286 Reject H0 0.005814  muC-muD -6.5 -12.02714 -0.97286 Reject H0 0.017701  >  > pairw.anova(data2$no\_customers,data2$teller,method = "dunnett", control= "A", conf.level = 0.95)  95% Dunnett confidence intervals  Diff Lower Upper Decision  muB-muA -7.093845 -11.682457 -2.505233 Reject H0  muC-muA -6.107966 -10.698823 -1.517109 Reject H0  muD-muA 0.40339 -4.394966 5.201746 FTR H0  #Another function for Multiple Comparisons with Tukey method   |  | | --- | | > TukeyHSD(anova\_model)  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = no\_customers ~ teller, data = data2)  $teller  diff lwr upr p adj  B-A -7.1 -12.173637 -2.026363 0.0046830  C-A -6.1 -11.173637 -1.026363 0.0154296  D-A 0.4 -4.899241 5.699241 0.9964410  C-B 1.0 -3.837523 5.837523 0.9355463  D-B 7.5 2.426363 12.573637 0.0028926  D-C 6.5 1.426363 11.573637 0.0096085  > a<-TukeyHSD(anova\_model)  > plot(a) | |  | | |  | | --- | | > | |     **TWO-WAY ANOVA**  > setwd("C:/Mihaela/R Scripts")  >  > eating <- read.csv("eating.csv", header=TRUE)  > eating  score gender breakfast  1 10 M HIGH  2 7 M HIGH  3 9 M HIGH  4 6 M HIGH  5 8 M HIGH  6 5 M LOW  7 4 M LOW  8 7 M LOW  9 4 M LOW  10 5 M LOW  11 5 F HIGH  12 4 F HIGH  13 6 F HIGH  14 3 F HIGH  15 2 F HIGH  16 3 F LOW  17 4 F LOW  18 5 F LOW  19 1 F LOW  20 2 F LOW  > with(eating, tapply(score, list(breakfast = breakfast, gender = gender), mean))  gender  breakfast F M  HIGH 4 8  LOW 3 5  > with(eating, tapply(score, list(breakfast = breakfast, gender = gender), sd))  gender  breakfast F M  HIGH 1.581139 1.581139  LOW 1.581139 1.224745  > with(eating, tapply(score, list(breakfast = breakfast, gender = gender), length))  gender  breakfast F M  HIGH 5 5  LOW 5 5   |  | | --- | | > ##Create new variable with the levels of the two factors  > eating$new <- with(eating, interaction(gender, breakfast), drop = TRUE)  > eating$new  [1] M.HIGH M.HIGH M.HIGH M.HIGH M.HIGH M.LOW M.LOW M.LOW M.LOW M.LOW F.HIGH F.HIGH F.HIGH  [14] F.HIGH F.HIGH F.LOW F.LOW F.LOW F.LOW F.LOW  Levels: F.HIGH M.HIGH F.LOW M.LOW  #To check for normality in each group: ﬁrst, let us take a look at some stripplots, one for each cell  > library(lattice)  > stripplot(~score | new, data = eating)    # The stripplots above are not very informative because there are only five observations in each cell  > plot(eating$score~ eating$gender)    > plot(eating$score~ eating$breakfast)    > plot(eating$score~interaction(eating$gender, eating$breakfast))      > attach(eating)  > ## Test of Normality  > by(score, list(breakfast = breakfast, gender = gender), shapiro.test)  breakfast: HIGH  gender: F  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.98676, p-value = 0.9672  ---------------------------------------------------------------------------  breakfast: LOW  gender: F  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.98676, p-value = 0.9672  ---------------------------------------------------------------------------  breakfast: HIGH  gender: M  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.98676, p-value = 0.9672  ---------------------------------------------------------------------------  breakfast: LOW  gender: M  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.83274, p-value = 0.1458  #The tests show that every cell passes its respective normality test at signiﬁcance level α = 0.05  > library(car)  #Test for homogeneity of Variances  > leveneTest(y=score, group=new)#test for homogeneity of variances  Levene's Test for Homogeneity of Variance (center = median)  Df F value Pr(>F)  group 3 0.2857 0.835  16  #Another way of writing the Test for homogeneity of Variances  > leveneTest(score ~ gender\*breakfast, data=eating)  Levene's Test for Homogeneity of Variance (center = median)  Df F value Pr(>F)  group 3 0.2857 0.835  16    > #TWO way ANOVA model  > anova\_model = aov(score~breakfast\*gender)  > qqnorm(anova\_model$res)  > plot(anova\_model$fitted,anova\_model$res,xlab="Fitted",ylab="Residuals")    > TukeyHSD(anova\_model, which = "gender")  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = score ~ breakfast + gender)  $gender  diff lwr upr p adj  M-F 3 1.534699 4.465301 0.000465  > a<-TukeyHSD(anova\_model, which = "gender")  > plot(a)      > TukeyHSD(anova\_model, which = "breakfast")  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = score ~ breakfast + gender)  $breakfast  diff lwr upr p adj  LOW-HIGH -2 -3.465301 -0.5346987 0.0104002  > b<-TukeyHSD(anova\_model, which = "breakfast")  > plot(b) | |  | |  |     > anova\_model = aov(score~breakfast\*gender)  > summary(anova\_model)  Df Sum Sq Mean Sq F value Pr(>F)  breakfast 1 20 20.00 8.889 0.008814 \*\*  gender 1 45 45.00 20.000 0.000385 \*\*\*  breakfast:gender 1 5 5.00 2.222 0.155487  Residuals 16 36 2.25 #within  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  > interaction.plot(eating$breakfast, eating$gender, eating$score)    **INTERPRETATION**  There appears to be significant main effects for both protein level (F=8.89 (1,16), p<.01) and gender (F=20.00 (1,16), p<.01). There was not a significant interaction effect (F=2.22 (1,16), not significant).  Based on this data, it appears that a high protein diet results in a better fitness test score. Additionally, young men seem to have a significantly higher fitness test score than women.  **EXAMPLE 4**  > exercise<-read.csv("exercise.csv", header=TRUE)  > exercise$type  [1] CONTROL CONTROL CONTROL CONTROL CONTROL ROLLER ROLLER ROLLER  [9] ROLLER ROLLER TRADITIONAL TRADITIONAL TRADITIONAL TRADITIONAL TRADITIONAL CONTROL  [17] CONTROL CONTROL CONTROL CONTROL ROLLER ROLLER ROLLER ROLLER  [25] ROLLER TRADITIONAL TRADITIONAL TRADITIONAL TRADITIONAL TRADITIONAL  Levels: CONTROL ROLLER TRADITIONAL  > table(exercise$gender, exercise$type)    CONTROL ROLLER TRADITIONAL  F 5 5 5  M 5 5 5  >  > exercise$new <- with(exercise, interaction(gender, type), drop = TRUE)  > exercise$new  [1] F.CONTROL F.CONTROL F.CONTROL F.CONTROL F.CONTROL F.ROLLER  [7] F.ROLLER F.ROLLER F.ROLLER F.ROLLER F.TRADITIONAL F.TRADITIONAL  [13] F.TRADITIONAL F.TRADITIONAL F.TRADITIONAL M.CONTROL M.CONTROL M.CONTROL  [19] M.CONTROL M.CONTROL M.ROLLER M.ROLLER M.ROLLER M.ROLLER  [25] M.ROLLER M.TRADITIONAL M.TRADITIONAL M.TRADITIONAL M.TRADITIONAL M.TRADITIONAL  Levels: F.CONTROL M.CONTROL F.ROLLER M.ROLLER F.TRADITIONAL M.TRADITIONAL  > plot(exercise$count~ exercise$gender)    > plot(exercise$count~ exercise$type)     |  | | --- | | > plot(exercise$count~interaction(exercise$gender, exercise$type)) | |  | | |  | | --- | |  | | |  | | |  | | |  | | | >Test of normality  > by(exercise$count, list(type = exercise$type, gender = exercise$gender), shapiro.test)  type: CONTROL  gender: F  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.76672, p-value = 0.0422  ---------------------------------------------------------------------------  type: ROLLER  gender: F  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.95235, p-value = 0.754  ---------------------------------------------------------------------------  type: TRADITIONAL  gender: F  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.99929, p-value = 0.9998  ---------------------------------------------------------------------------  type: CONTROL  gender: M  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.93855, p-value = 0.6557  ---------------------------------------------------------------------------  type: ROLLER  gender: M  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.88783, p-value = 0.3463  ---------------------------------------------------------------------------  type: TRADITIONAL  gender: M  Shapiro-Wilk normality test  data: dd[x, ]  W = 0.93855, p-value = 0.6557 | | |   > Test of homogeneity  > bartlett.test(exercise$count ~ exercise$new)  Bartlett test of homogeneity of variances  data: exercise$count by exercise$new  Bartlett's K-squared = 2.3432, df = 5, p-value = 0.7999  > par(mfrow=c(2,2))  > plot(exercise.m1)  > par(mfrow=c(1,1))     |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | > exercise.m2 <- aov(count ~ gender \* type, exercise)  > summary(exercise.m2)  Df Sum Sq Mean Sq F value Pr(>F)  gender 1 0.03 0.03 0.006 0.941  type 2 211.27 105.63 17.904 1.74e-05 \*\*\*  gender:type 2 3.27 1.63 0.277 0.761  Residuals 24 141.60 5.90  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  > exercise.m1 <- aov(count ~ gender + type, exercise)  > summary(exercise.m1)  Df Sum Sq Mean Sq F value Pr(>F)  gender 1 0.03 0.03 0.006 0.939  type 2 211.27 105.63 18.959 8.35e-06 \*\*\*  Residuals 26 144.87 5.57  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1   |  | | --- | | > | | interaction.plot(exercise$gender, exercise$type, exercise$count) | |  | | |  | | --- | |  | | | |  | |  | |
| |  | | --- | |  | |  |   > TukeyHSD(exercise.m1, which = "type")  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = count ~ gender + type, data = exercise)  $type  diff lwr upr p adj  ROLLER-CONTROL 3.3 0.6768665 5.923133 0.0116382  TRADITIONAL-CONTROL 6.5 3.8768665 9.123133 0.0000048  TRADITIONAL-ROLLER 3.2 0.5768665 5.823133 0.0145754  > b<-TukeyHSD(exercise.m1, which = "type")  > plot(b)    INTERPRETATION   1. The main effect of exercise type says that sit-up ability is increased the most with traditional sit-ups, less with the Roller machine, and the least in the control group. 2. The lack of a main effect of gender means that males and females were equivalent in sit-up performance. 3. The lack of an interaction means that the effect of exercise type does not depend on gender. |