> #################################

> # R session 2.2: Extra exercise #

> #################################

>

> # Use the function setwd to set the working directory to

> # the folder where you stored the file tbad.txt

> setwd("C:\\Users\\lucp2169\\Documents\\FLAMES\\Cursussen\\Basic parametric statistics\\versie 2018\\Session 1")

>

> # Select the file "TBAD.txt" and check the variables

> tbad<-read.table("tbad.txt",header=T,sep="\t",dec=".")

> str(tbad)

'data.frame': 186 obs. of 9 variables:

$ Specialty : Factor w/ 6 levels "fp","gp","im",..: 6 6 6 6 3 5 3 1 6 6 ...

$ Secondary.specialty : int 0 1 0 0 0 0 0 0 0 0 ...

$ Certification.level : int 1 1 2 1 1 0 0 1 1 1 ...

$ Gender : int 0 0 0 1 0 0 0 0 0 1 ...

$ Medical.school : int 0 1 1 1 1 1 1 0 0 1 ...

$ Residence : int 0 0 0 0 0 1 1 1 1 0 ...

$ Years.of.experience : int 37 35 0 21 8 0 38 22 23 0 ...

$ Total.average.costs.per.patient.per.month: num 47.6 25.9 53.7 74.1 18.2 ...

$ Total.patients.per.month : int 2108 18 430 255 12 707 71 88...

>

> # Secondary.specialty, Certification.level, Gender,

> # Medical.school and Residence should be factors

> # and add labels

> tbad$Secondary.specialty<-factor(tbad$Secondary.specialty,

+ levels=0:1,labels=c("no","yes"))

> tbad$Certification.level<-factor(tbad$Certification.level)

> tbad$Gender=factor(tbad$Gender,

+ levels=0:1,labels=c("man","woman"))

> tbad$Medical.school<-factor(tbad$Medical.school,

+ levels=0:1,labels=c("USA","Foreign"))

> tbad$Residence<-factor(tbad$Residence,

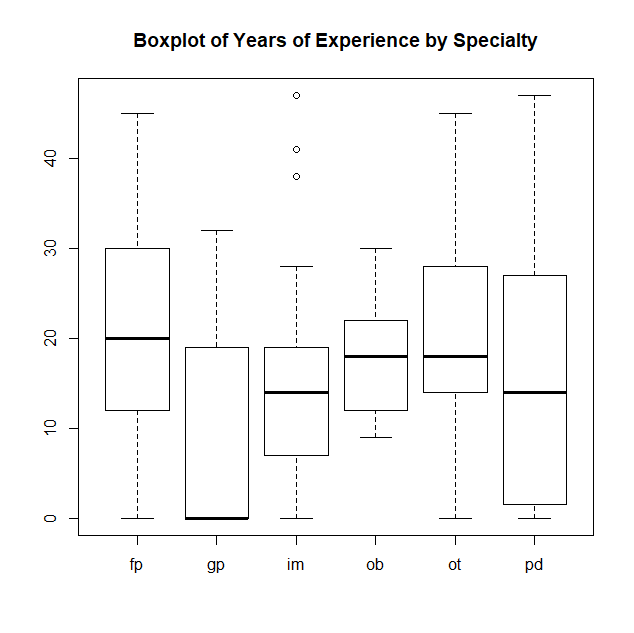
+ levels=0:1,labels=c("USA","Foreign"))

>

> # 1-way anova experience by specialties

> boxplot(tbad$Years.of.experience~tbad$Specialty,

+ main="Boxplot of Years of Experience by Specialty")



> lmanova <- lm(Years.of.experience~Specialty,

+ data=tbad)

> anova(lmanova)

Analysis of Variance Table

Response: Years.of.experience

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

Specialty 5 2177.2 435.44 2.6559 0.02415 \*

Residuals 180 29510.7 163.95

---

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

> pairwise.t.test(tbad$Years.of.experience, tbad$Specialty,

+ p.adj="none")

Pairwise comparisons using t tests with pooled SD

data: tbad$Years.of.experience and tbad$Specialty

fp gp im ob ot

gp 0.00058 - - - -

im 0.07583 0.05238 - - -

ob 0.49669 0.05077 0.60778 - -

ot 0.75260 0.01986 0.36798 0.76346 -

pd 0.10347 0.01952 0.73683 0.75484 0.47551

P value adjustment method: none

> pairwise.t.test(tbad$Years.of.experience, tbad$Specialty,

+ p.adj="bonferroni")

Pairwise comparisons using t tests with pooled SD

data: tbad$Years.of.experience and tbad$Specialty

fp gp im ob ot

gp 0.0087 - - - -

im 1.0000 0.7857 - - -

ob 1.0000 0.7615 1.0000 - -

ot 1.0000 0.2979 1.0000 1.0000 -

pd 1.0000 0.2928 1.0000 1.0000 1.0000

P value adjustment method: bonferroni

> aovanova <- aov(Years.of.experience ~ Specialty, data=tbad)

> TukeyHSD(aovanova)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = Years.of.experience ~ Specialty, data = tbad)

$`Specialty`

diff lwr upr p adj

gp-fp -12.0884211 -22.028955 -2.147887 0.0075414

im-fp -4.9977778 -13.060096 3.064541 0.4775713

ob-fp -2.8033333 -14.660125 9.053459 0.9838477

ot-fp -1.2584615 -12.741599 10.224676 0.9995764

pd-fp -4.0771429 -11.253778 3.099493 0.5755135

im-gp 7.0906433 -3.368600 17.549887 0.3737272

ob-gp 9.2850877 -4.315585 22.885761 0.3656686

ot-gp 10.8299595 -2.446231 24.106150 0.1799561

pd-gp 8.0112782 -1.781520 17.804076 0.1774317

ob-im 2.1944444 -10.100476 14.489365 0.9955765

ot-im 3.7393162 -8.195675 15.674307 0.9454391

pd-im 0.9206349 -6.958810 8.800080 0.9994211

ot-ob 1.5448718 -13.220850 16.310594 0.9996624

pd-ob -1.2738095 -13.007018 10.459399 0.9995955

pd-ot -2.8186813 -14.174170 8.536807 0.9799283

> library(DTK)

> TK.test(tbad$Years.of.experience, tbad$Specialty)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = x ~ f)

$`f`

diff lwr upr p adj

gp-fp -12.0884211 -22.028955 -2.147887 0.0075414

im-fp -4.9977778 -13.060096 3.064541 0.4775713

ob-fp -2.8033333 -14.660125 9.053459 0.9838477

ot-fp -1.2584615 -12.741599 10.224676 0.9995764

pd-fp -4.0771429 -11.253778 3.099493 0.5755135

im-gp 7.0906433 -3.368600 17.549887 0.3737272

ob-gp 9.2850877 -4.315585 22.885761 0.3656686

ot-gp 10.8299595 -2.446231 24.106150 0.1799561

pd-gp 8.0112782 -1.781520 17.804076 0.1774317

ob-im 2.1944444 -10.100476 14.489365 0.9955765

ot-im 3.7393162 -8.195675 15.674307 0.9454391

pd-im 0.9206349 -6.958810 8.800080 0.9994211

ot-ob 1.5448718 -13.220850 16.310594 0.9996624

pd-ob -1.2738095 -13.007018 10.459399 0.9995955

pd-ot -2.8186813 -14.174170 8.536807 0.9799283

> DTK.test(tbad$Years.of.experience, tbad$Specialty)

[[1]]

[1] 0.05

[[2]]

Diff Lower CI Upper CI

gp-fp -12.0884211 -23.355176 -0.821666

im-fp -4.9977778 -13.287112 3.291557

ob-fp -2.8033333 -11.026749 5.420082

ot-fp -1.2584615 -14.970565 12.453642

pd-fp -4.0771429 -12.030240 3.875954

im-gp 7.0906433 -4.096221 18.277508

ob-gp 9.2850877 -1.857173 20.427349

ot-gp 10.8299595 -4.477256 26.137175

pd-gp 8.0112782 -2.949970 18.972526

ob-im 2.1944444 -6.854030 11.242919

ot-im 3.7393162 -10.722185 18.200817

pd-im 0.9206349 -7.868723 9.709993

ot-ob 1.5448718 -13.238365 16.328108

pd-ob -1.2738095 -10.216184 7.668565

pd-ot -2.8186813 -15.934825 10.297463

>

>

> # 2-way anova experience by specialties and gender

> lmanova <- lm(Years.of.experience~Specialty+Gender+Specialty\*Gender,

+ data=tbad)

> anova(lmanova)

Analysis of Variance Table

Response: Years.of.experience

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

Specialty 5 2177.2 435.44 2.6477 0.02465 \*

Gender 1 230.5 230.47 1.4014 0.23810

Specialty:Gender 5 664.4 132.89 0.8080 0.54537

Residuals 174 28615.8 164.46

---

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

> lmanova <- lm(Years.of.experience~Specialty+Gender,

+ data=tbad)

> anova(lmanova)

Analysis of Variance Table

Response: Years.of.experience

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

Specialty 5 2177.2 435.44 2.662 0.0239 \*

Gender 1 230.5 230.47 1.409 0.2368

Residuals 179 29280.3 163.58

---

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

>

> # two-way anova based on health.txt

> health<-read.table("Health.txt",

+ header=T,sep=",")

> str(health)

'data.frame': 562 obs. of 6 variables:

$ exercise : int 130 146 160 158 165 132 135 159 149 168 ...

$ age : int 49 39 32 33 38 45 41 38 38 38 ...

$ gender : Factor w/ 2 levels "female","male": 1 1 1 1 1 2 1 1 1 1 ...

$ bmi : Factor w/ 3 levels "high","low","normal": 3 3 3 3 3 3 3 3 3 3 ...

$ profession: Factor w/ 3 levels "none","sitting",..: 2 3 2 3 3 2 3 3 1 2 ...

$ children : Factor w/ 2 levels "none","one or more": 2 1 2 2 2 1 1 2 1 2 ...

>

> lm1=lm(exercise ~ gender+bmi+gender:bmi,data=health)

> anova(lm1)

Analysis of Variance Table

Response: exercise

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

gender 1 331 330.68 1.9416 0.1640

bmi 2 54 26.83 0.1575 0.8543

gender:bmi 2 84 42.00 0.2466 0.7815

Residuals 556 94694 170.31

> lm1a=lm(exercise ~ gender+bmi,data=health)

> anova(lm1a)

Analysis of Variance Table

Response: exercise

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

gender 1 331 330.68 1.9469 0.1635

bmi 2 54 26.83 0.1579 0.8539

Residuals 558 94778 169.85

> lm1b=lm(exercise ~ gender,data=health)

> anova(lm1b)

Analysis of Variance Table

Response: exercise

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

gender 1 331 330.68 1.9528 0.1628

Residuals 560 94831 169.34

>

> lm2=lm(exercise ~ gender+profession+gender:profession,data=health)

> anova(lm2)

Analysis of Variance Table

Response: exercise

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

gender 1 331 330.7 2.4630 0.1171237

profession 2 18209 9104.3 67.8113 < 2.2e-16 \*\*\*

gender:profession 2 1974 987.0 7.3517 0.0007058 \*\*\*

Residuals 556 74649 134.3

---

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

>

> lm3=lm(exercise ~ bmi+profession+bmi:profession,data=health)

> anova(lm3)

Analysis of Variance Table

Response: exercise

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

bmi 2 43 21.5 0.1543 0.8571

profession 2 17903 8951.3 64.2964 <2e-16 \*\*\*

bmi:profession 4 228 57.1 0.4103 0.8013

Residuals 553 76988 139.2

---

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

> lm3a=lm(exercise ~ bmi+profession,data=health)

> anova(lm3a)

Analysis of Variance Table

Response: exercise

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

bmi 2 43 21.5 0.1549 0.8565

profession 2 17903 8951.3 64.5698 <2e-16 \*\*\*

Residuals 557 77217 138.6

> lm3b=lm(exercise ~ profession,data=health)

> anova(lm3b)

Analysis of Variance Table

Response: exercise

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

profession 2 17917 8958.4 64.829 < 2.2e-16 \*\*\*

Residuals 559 77245 138.2

>

> lm4=lm(exercise ~ bmi+children+bmi:children,data=health)

> anova(lm4)

Analysis of Variance Table

Response: exercise

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

bmi 2 43 21.48 0.1272 0.88062

children 1 361 361.35 2.1396 0.14410

bmi:children 2 858 429.07 2.5406 0.07973 .

Residuals 556 93900 168.88

> lm4a=lm(exercise ~ bmi+children,data=health)

> anova(lm4a)

Analysis of Variance Table

Response: exercise

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

bmi 2 43 21.48 0.1265 0.8812

children 1 361 361.35 2.1279 0.1452

Residuals 558 94758 169.82

> lm4b=lm(exercise ~ children,data=health)

> anova(lm4b)

Analysis of Variance Table

Response: exercise

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

children 1 368 368.00 2.174 0.1409

Residuals 560 94794 169.28

>

> lm5=lm(exercise ~ profession+children+profession:children,data=health)

> anova(lm5)

Analysis of Variance Table

Response: exercise

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

profession 2 17917 8958.4 65.3310 < 2e-16 \*\*\*

children 1 585 585.4 4.2694 0.03927 \*

profession:children 2 420 209.8 1.5304 0.21737

Residuals 556 76240 137.1

> lm5a=lm(exercise ~ profession+children,data=health)

> anova(lm5a)

Analysis of Variance Table

Response: exercise

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

profession 2 17917 8958.4 65.2071 < 2e-16 \*\*\*

children 1 585 585.4 4.2613 0.03945 \*

Residuals 558 76660 137.4

>