# Partial Script of some commands used to produce useful output

# for Week 2 R assignment Juul Descriptive Data Analysis

R version 2.12.2 (2011-02-25)

Copyright (C) 2011 The R Foundation for Statistical Computing

ISBN 3-900051-07-0

Platform: i386-pc-mingw32/i386 (32-bit)

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

# load ISwR package

> local({pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE)

+ if(nchar(pkg)) library(pkg, character.only=TRUE)})

> attach(juul)

>

> summary(juul)

> juul[0]

> detach(juul)

# make explicit and recode categorical variables - they are not numeric!

> juul$sex <- factor(juul$sex,labels=c("M","F"))

> juul$menarche <- factor(juul$menarche,labels=c("No","Yes"))

> juul$tanner <- factor(juul$tanner,

+ labels=c("I","II","III","IV","V"))

> juul[0]

> juul$tanner

[1] <NA> <NA> <NA> <NA> <NA> I I I I I I I I I

[15] I I I I I I I I I I I I I I

[29] I I I I I I I I I I I I I I

[43] I I I I I I I I I I I I I I

[57] I I I I I I I I I I I I I I

[71] I I I I I I I I I I I I I I

[85] I I I I I <NA> I I I I I I I I

.......

[1317] <NA> V <NA> V <NA> <NA> <NA> <NA> V V <NA> <NA> V <NA>

[1331] <NA> <NA> <NA> V V <NA> <NA> <NA> <NA>

Levels: I II III IV V

> summary(juul)

age menarche sex igf1 tanner

Min. : 0.170 No :369 M :621 Min. : 25.0 I :515

1st Qu.: 9.053 Yes :335 F :713 1st Qu.:202.2 II :103

Median :12.560 NA's:635 NA's: 5 Median :313.5 III : 72

Mean :15.095 Mean :340.2 IV : 81

3rd Qu.:16.855 3rd Qu.:462.8 V :328

Max. :83.000 Max. :915.0 NA's:240

NA's : 5.000 NA's :321.0

testvol

Min. : 1.000

1st Qu.: 1.000

Median : 3.000

Mean : 7.896

3rd Qu.: 15.000

Max. : 30.000

NA's :859.000

hist(age,ylab="Frequency",xlab="Age",main="Age Distribution")

> qqplot(juul$igf1)

> qqnorm(juul$igf1, main="Insulin Growth Factor Data Distribution")

> qqline(juul$igf1)

> plot(juul$sex,juul$tanner)

> plot(juul$sex,juul$tanner,ylab="Tanner Stage",xlab="Gender",main="Tanner Stage by Gender")

> tapply(igf1, tanner, mean, na.rm=T) #gives mean for each tanner group (ignores missing values)

1 2 3 4 5

207.4727 352.6714 483.2222 513.0172 465.3344

> aggregate(juul[c("age","igf1")],list(sex=juul$sex), mean, na.rm=T)

sex age igf1

1 M 15.38436 310.8866

2 F 14.84363 368.1006

> plot(igf1~age)

> plot(testvol~sex)

> hist(testvol)

> plot(testvol~age)

> # here i stopped - i was looking for variation of testvol by age.

# commands below were my experiments with age, how i learned about qqnorm and qqline,

# and tapply

> mid.age <- c(2.5,7.5,13,16.5,17.5,19,22.5,44.5,70.5)

> acc.count <- c(28,46,58,20,31,64,149,316,103)

> age.acc <- rep(mid.age,acc.count)

> age.acc

[1] 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5

[16] 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 7.5 7.5

[31] 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5

[46] 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5

[61] 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 13.0

[76] 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0

[91] 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0 13.0

....

[781] 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5

[796] 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5 70.5

[811] 70.5 70.5 70.5 70.5 70.5

> brk <- c(0,5,10,16,17,18,20,25,60,80)

> hist(age.acc,breaks=brk)

> plot(testvol~age.acc)

> x[0]

numeric(0)

> n <- length(x)

> n

[1] 50

> sort(x)

[1] -2.57055422 -1.94929248 -1.64849369 -1.58801690 -1.38463293 -1.24241335

[7] -1.15424037 -1.11832765 -1.11325427 -1.08845786 -0.91218597 -0.77607024

[13] -0.76630927 -0.72641733 -0.72567415 -0.57692246 -0.54955181 -0.48058815

[19] -0.47962710 -0.47685889 -0.46457077 -0.44342369 -0.34999468 -0.34465189

[25] -0.30519139 -0.25912802 -0.20553430 -0.16576352 -0.11342019 -0.05246701

[31] 0.00962627 0.01110214 0.19306541 0.29081506 0.29762755 0.33202681

[37] 0.33941154 0.37256749 0.41819657 0.52116920 0.56235030 0.76900611

[43] 0.80304833 0.81404361 0.83499934 1.29929015 1.43623704 1.44957178

[49] 1.47568367 1.48846897

> plot(sort(x),(1:n)/n,type="s",ylim=c(0,1))

> qqnorm(x)

> qqline(x)

> par(mfrow=c(1,2))

> boxplot(IgM)

> boxplot(log(IgM))

>

> attach(red.cell.folate)

> par(mfrow=c(1,1))

> tapply(folate,ventilation,mean)

N2O+O2,24h N2O+O2,op O2,24h

316.6250 256.4444 278.0000

> help(tapply)

> groups <- as.factor(rbinom(32, n = 5, prob = 0.4))

> groups

[1] 14 11 13 14 11

Levels: 11 13 14

> tapply(groups, groups, length)

11 13 14

2 1 2

> table(groups)

groups

11 13 14

2 1 2

>