Statistics with R – Intermediate Level

Section 2

Mean Difference

Lesson 5 - Independent-Sample T Test

```
sp = read.csv("spanish.csv")

View(sp)

#########
### how to perform the independent sample t test
#########

### we will determine whether there is a significant
difference in average grade
### between the students who took the Spanish course and
those who did not take it

#########
### Basic assumptions:

# the dependent variable is normally distributed in both
groups
# the dependent variable does not present important
outliers in any group
# the group variances are equal*
```

```
### we are going to check only the assumptions marked with
an asterisk (*)
##########

### to check the assumption of homogeneity of variances
### we will use the leveneTest function in the car package

### load the package

require(car)

leveneTest(sp$score, sp$course)

### to compute the t test, we use the t.test function

### if the group variances are equal

t.test(sp$score~sp$course, var.equal=T)

### if the group variances are NOT equal we will execute
the Welch test

t.test(sp$score~sp$course, var.equal=F)
```

Lesson 6 - Paired-Sample T Test

```
mat = read.csv("math.csv")

View(mat)

#########
### how to perform a paired sample t test
#########

### we will check whether there is a significant difference
between
### the average grades of the tests, before and after the
math course

#########
### Basic assumptions:
```

```
# the differences between the dependent variables are
normally distributed
# the differences between the dependent variables do not
present important outliers
##########

### run the t.test function

t.test(mat$grade2, mat$grade1, paired=T)
```

Lesson 7 - Oneway ANOVA

```
vit = read.csv("vitamin1.csv")

View(vit)

#########
### how to perform the one-way analysis of variance
#########

### we will check whether there is a difference between the
three groups
### (placebo, low dose, high dose) with respect to the
average effort resistance

#########
### Basic assumptions:

# the dependent variable is normally distributed in all
groups
# the dependent variable does not present important
outliers in any group
# the group variances are equal*

### we are going to check only the assumptions marked with
an asterisk (*)
##########

### to check the assumption of equality of variances

require(car)

leveneTest(vit$effort, vit$dose)
```

```
##########

### run the one-way ANOVA using the aov function

### if the group variances are equal

aov1 = aov(effort~dose, data=vit)
summary(aov1)

### if the group variances are NOT equal, we run the Welch test

oneway.test(effort~dose, data=vit, var.equal=F)

##########

### how to perform the simple (post-hoc) comparisons

### Tukey HSD for equal variances

TukeyHSD(aov1)    # you must get aov1 first

### Bonferroni, also for equal variances

pairwise.t.test(vit$effort, vit$dose, p.adjust.method =
"bonferroni")

### post-hoc tests for unequal variances do not seem to be available in R :(
```

Lesson 8 - Twoway ANOVA – Basics

```
vit = read.csv("vitamin2.csv")

View(vit)

#########
### how to perform the two-way analysis of variance
#########

### we will determine whether the factors dose of vitamin and gender
```

```
### influence the employees' effort resistance

#########
### Basic assumptions:

# the dependent variable is normally distributed in all
groups
# the dependent variable does not present important
outliers in any group
# the group variances are equal
##########

### to check the equality of variances, we must create a
### separate variable to define the six groups
### placebo-male, low dose-male, high dose-male
### placebo-female, low dose-female, high dose-female
### (hint: you can do that with brackets or with the
revalue function in the plyr package)
### then run the leveneTest function in the car package

### perform the two way ANOVA using the aov function

aov1 = aov(effort~dose+gender+dose*gender, data=vit)
summary(aov1)

### since the interaction effect is statistically
significant,
### we are going to compute the simple main effects of the
two factors
```

Lesson 9 - Twoway ANOVA - Simple Main Effects

```
vit = read.csv("vitamin2.csv")

View(vit)

#########
### how to compute the simple main effects
#########

### the simple main effect of a factor is the effect of
that factor
### at every level of the other factor
```

```
##########

### the simple main effect of the factor gender is
### the effect of gender on the effort resistance for each
dose group:
### placebo, low dose and high dose

### we are going to build a separate data frame for each
dose group
### with the help of the brackets

### dose = placebo

vitp = vit[vit$dose=="placebo",]

View(vitp)

### run the ANOVA

aov1 = aov(effort~gender, data=vitp)
summary(aov1)

### perform the Tukey HSD (simple comparison)

TukeyHSD(aov1)


### dose = low

vitld = vit[vit$dose=="low dose",]

View(vitld)

### run the ANOVA

aov2 = aov(effort~gender, data=vitld)
summary(aov2)

### get the Tukey HSD

TukeyHSD(aov2)
```

```
### dose = high

vithd = vit[vit$dose=="high dose",]

View(vithd)

### run the ANOVA

aov3 = aov(effort~gender, data=vithd)
summary(aov3)

### get the Tukey HSD

TukeyHSD(aov3)

##########

### the simple main effect of the factor dose is the
### effect of dose on the effort resistance for each gender
group:
### male and female

### we are going to build a separate data frame for each
gender group
### using the brackets

### gender = male

vitm = vit[vit$gender=="male",]

View(vitm)

### run the ANOVA

aov4 = aov(effort~dose, data=vitm)
summary(aov4)

### perform Tukey HSD

TukeyHSD(aov4)

### gender = female

vitf = vit[vit$gender=="female",]
```

```
View(vitf)

### run the ANOVA

aov5 = aov(effort~dose, data=vitf)
summary(aov5)

### perform Tukey HSD

TukeyHSD(aov5)
```

Lesson 10 - Threeway ANOVA – Basics

```
vit = read.csv("vitamin3.csv")

View(vit)

#########
### how to perform the three-way analysis of variance
#########

### we will determine whether the factors dose of vitamin, 
gender and type
### influence the employees' effort resistance

#########
### Basic assumptions:

# the dependent variable is normally distributed in all 
groups
# the dependent variable does not present important 
outliers in any group
# the group variances are equal
##########

### to check the equality of variances, we must create a
### separate variable to define the twelve groups
### placebo-male-blue collar, low dose-male-blue collar, 
high dose-male-blue collar
### placebo-male-white collar, low dose-male-white collar, 
high dose-male-white collar
###... and so on
```

```
### then run the leveneTest function in the car package

### perform the three way ANOVA using the aov function

aov1 =
aov(effort~dose+gender+type+dose*gender+dose*type+gender*ty
pe+dose*gender*type, data=vit)
summary(aov1)

### since the third order interaction effect is
statistically significant
### we must compute the simple second order interaction
effects
```

Lesson 11 - Threeway ANOVA - Simple Second Order Interaction Effects

```
vit = read.csv("vitamin3.csv")

View(vit)

#########
### how to compute the simple second order interaction
effects
#########

### the simple second order interaction effect is
### the interaction effect of two factors at each level of
the third factor

### here we have three simple second order interaction
effects
### dose * gender for each type of employee
### dose * type of employee for each gender category
### gender * type of employee for each dose of vitamin

#############

### we will compute the simple second order interaction
effect
### of dose and type of employee for each gender category:
### male and female
```

```
### gender = male

### create a new data frame with the male subjects

vitm <- vit[vit$gender=="male",]

aov1 <- aov(effort~dose*type, data=vitm)
summary(aov1)

### gender = female

### create a new data frame with the female subjects

vitf <- vit[vit$gender=="female",]

aov2 <- aov(effort~dose*type, data=vitf)
summary(aov2)

### since both second order interactions are statistically
significant
### we will compute the simple main effects for both
variables
### dose and type of employee
```

Lesson 12 - Threeway ANOVA - Simple Main Effects

```
vit = read.csv("vitamin3.csv")

View(vit)

#########
### how to compute the simple main effects
#########

### in a three-way ANOVA, the simple main effect of a
factor is the effect of that factor
### at every combination of the levels of the other two
factors

### we will compute the simple main effecs of the factor
dose
### at each combination of gender and type of employee:
### male-blue collar, female-blue collar,
```

```
### male-white collar, female-white collar

######## male - blue collar

### create a new data frame with the male subjects, blue
collar

vitmbcol <- vit[vit$gender=="male" & vit$type=="blue
collar",]

### run the ANOVA

aov1 <- aov(effort~dose, data=vitmbcol)
summary(aov1)

### compute the Tukey HSD

TukeyHSD(aov1)

####### female - blue collar

### create a new data frame with the female subjects, blue
collar

vitfbcol <- vit[vit$gender=="female" & vit$type=="blue
collar",]

### run the ANOVA

aov2 <- aov(effort~dose, data=vitfbcol)
summary(aov2)

### compute the Tukey HSD

TukeyHSD(aov2)

########## male - white collar

### create a new data frame with the male subjects, white
collar

vitmwcol <- vit[vit$gender=="male" & vit$type=="white
collar",]
```

```
### run the ANOVA

aov3 <- aov(effort~dose, data=vitmwcol)
summary(aov3)

### compute the Tukey HSD

TukeyHSD(aov3)

############# female - white collar

### create a new data frame with the female subjects, white
collar

vitfwcol <- vit[vit$gender=="female" & vit$type=="white
collar",]

### run the ANOVA

aov4 <- aov(effort~dose, data=vitfwcol)
summary(aov4)

### compute the Tukey HSD

TukeyHSD(aov4)
```

Lesson 13 - Oneway MANOVA

```
vit = read.csv("vitamin-m.csv")

View(vit)

#########
### how to perform the one-way multivariate analysis of
variance (MANOVA)
#########

### we will check whether there is a difference between the
three groups
### (placebo, low dose, high dose) with respect to the
average effort resistance
### and the average stress resistance
```

```
#########
### Basic assumptions:

# the dependent variables are normally distributed in all
factor groups
# the dependent variables do not present important outliers
in any group
# the relationship between the dependent variables are
approximately linear in all groups
# the dependent variables are not strongly intercorrelated
(there is no multicollinearity)
# the group variances are equal for all the dependent
variables
# the group covariances are equal*

### we will check only the assumptions marked with an
asterisk (*)
##########

### check the assumption of homogeneity of covariances
first

### load the package biotools

require(biotools)

### create a new data frame with the dependent variables
only

vit2 <- vit[c(1,2)]

View(vit2)

### apply the Box's M test from biotools

boxM(vit2, vit$dose)

### before running MANOVA, we create a matrix with the
dependent variables
### this will be the dependent variable in the manova
function

y <- cbind(vit$effort, vit$stress)
```

```
print(y)

### create the factor dose

dose <- vit$dose

print(dose)

### run the manova function from the stats package

model <- manova(y~dose)
summary(model, test="Wilks")

### we used the Wilks test because the covariances are
equal

### other test options  Pillai, Hotelling, Roy

### the Pillai test (Pillai's trace) is recommended if the
covariances are not equal

summary(model, test="Pillai")

### N.B. Tukey HSD does NOT work with the manova function
### if you need it, you must use it for each dependent
variable separately
### i.e. you must run an univariate one-way ANOVA for each
dependent (using the aov function),
### and then apply the TukeyHSD function to the results
```

Lesson 14 - Mann-Whitney Test

```
sp = read.csv("spanish.csv")

View(sp)

#########
### how to run the Mann-Whitney test
#########

### we will check whether there is a difference between the
two groups of students
```

```
### with respect to the average test score, using the Mann-
Whitney test

wilcox.test(sp$score~sp$course)

### or without continuity correction

wilcox.test(sp$score~sp$course, correct=F)

### to get the group medians

require(psych)

describeBy(sp$score, sp$course)
```

Lesson 15 - Wilcoxon Test

```
mat = read.csv("math.csv")

View(mat)

#########
### how to run the Wilcoxon test
#########

### we are going to find out whether there is a significant
difference
### between the two grades (before and after the math
course) on average

wilcox.test(mat$grade2, mat$grade1, paired=T)

### or, without the continuity correction

wilcox.test(mat$grade2, mat$grade1, paired=T, correct=F)

### compute the medians

median(mat$grade2)

median(mat$grade1)
```

Lesson 16 - Kruskal-Wallis Test

```
vit = read.csv("vitamin1.csv")

View(vit)

#########
### how to run the Kruskall-Wallis test
#########

### we are going to check whether there is a significant
difference
### between the three groups with respect to the average
effort resistance

### run the test

kruskal.test(vit$effort~vit$dose)

### get the medians

require(psych)

describeBy(vit$effort, vit$dose)
```