Questions

(All answers for questions A and B are in the slides).

A. You want to empirically test whether LaPraire has premium products compared to Vichy. To answer the question you look up comparable products from a retail website and find the prices which are stored in the file called "prices.csv".

1. Read data in

2. Explore Descriptive Statistics for each variable including mean, median, SD and Var.

3. Plot histograms from each variable and superimpose a normal curve.

4. Change the colour of the histogram to red

5. Plot two histograms side by side

6. Conduct a t-test to test whether the prices of one company are significantly different from the other.

7. Repeat the analysis but use a non-parametric test

8. Combine the data into a data.frame and run the analysis again

9. Is there an obvious link between the prices of the two companies? If LaPraire releases a new product that costs £80, what would you predict Vichy to price their competing product?

B You are now interested in the product range from a third company "Clarins"

1 Look at box plots and descriptive statistics for all three companies.

2. What would you expect in terms of differences?

3. Is there an overall difference across the three groups?

4. Report the results from your test.

5. Where is the effect coming from? What groups differ from each other?

6. Are all assumptions of homogeneity of variance met?

C. Create a data.frame from the following data:

Sand	Silt	Pebbles	Glass
1.45	1.24	2.24	1.18
0.76	1.93	3.71	0.59
1.11	1.96	2.92	0.52
1.71	2.20	3.01	-0.74
0.97	3.93	6.33	-0.99

The data shows the influence of substrate type on growth rates of algae.

```
growth<- c(1.45, 0.76, 1.11, 1.71, 0.97, 1.24, 1.93, 1.96,
2.20, 3.93, 2.24, 3.71, 2.92, 3.01, 6.33, 1.18, 0.59, 0.52,
-0.74, -0.99)
```

```
substrate<-c('sand', 'sand', 'sand', 'sand', 'sand', 'silt',
'silt', 'silt', 'silt', 'silt', 'pebbles', 'pebbles',
'pebbles', 'pebbles', 'pebbles', 'glass', 'glass', 'glass',
'glass', 'glass')
```

```
data<- data.frame('growth'=growth, 'substrate'=substrate)
```

4. Explore your data

```
dim(data)

class(data)

str(data)

names(data)

is.data.frame(data)
```

5. Generate a box plot and look at your data

```
boxplot(growth~substrate, ylab='Algal Growth Rate (/
day)',xlab='Substrate Type')
```

6. Conduct a one-way ANOVA

```
results<- aov(growth~substrate, data=data)
```

7. Report the results from your ANOVA

```
summary(results)
```

8. Perform a post hoc test. Which levels differ from each other?

```
TukeyHSD(results)
```

D. Read csv file which contains a hypothetical sample of 27 participants who are divided into three stress reduction

treatment groups (mental, physical, and medical) and three age groups (young, mid, and old). Data file name “2_way_data.csv”.

1. Read data into an object called “data”.

   ```
   data<-read.csv(file=”2_way_data.csv”)
   ```

2. Explore your data

   ```
   dim(data)

   class(data)

   names(data)

   str(data)
   ```

3. Perform an ANOVA test and report each main effect and interaction.

   ```
   Results<-anova(lm(StressReduction ~ Treatment * Age, data))
   Results
   #You can use aov in which case you need to call summary(results)
   ```

4. Use either t-test or Tukeys post hoc test to test pairwise comparisons

   ```
   pairwise.t.test(data$StressReduction, data$Treatment, p.adj = "none")
   ```

5. Plot the interaction.