Exercise 2 - solution

Paul Blanche

Exercise A

Question 1

Research hypothesis: the average protein level of the milk at 6 weeks after calving of cows fed with lupins only is **larger** than that of cows fed with barley and lupins.

Null and alternative hypotheses:

- H_0: the average protein level of the milk at 6 weeks after calving of cows fed with lupins only is **equal** to than that of cows fed with barley and lupins.

- H_1: the average protein level of the milk at 6 weeks after calving of cows fed with lupins only is **different** to than that of cows fed with barley and lupins.

Statistical methods: we will compare the mean of the two groups using Welch's t-test. We plan to report the estimated mean in each group and their difference, the corresponding 95% confidence intervals and the p-value of the Welch's t-test.

**Prepare and load the data:** (similar to what it is done in the R-demo of Lecture 2).

```r
library(nlme)
# first keep only specific rows
d <- Milk[which(Milk$Time==6 & Milk$Diet %in% c("lupins","barley+lupins")),]
# second keep only specific columns
d <- d[,c("protein","Diet")]
# Finally we delete the levels no longer present in the data.
d$Diet <- droplevels(d$Diet)
summary(d)
```

```
##     protein                Diet
##  Min.   :2.720   barley+lupins:27
##  1st Qu.:3.118   lupins       :27
##  Median :3.300
##  Mean   :3.336
##  3rd Qu.:3.547
```
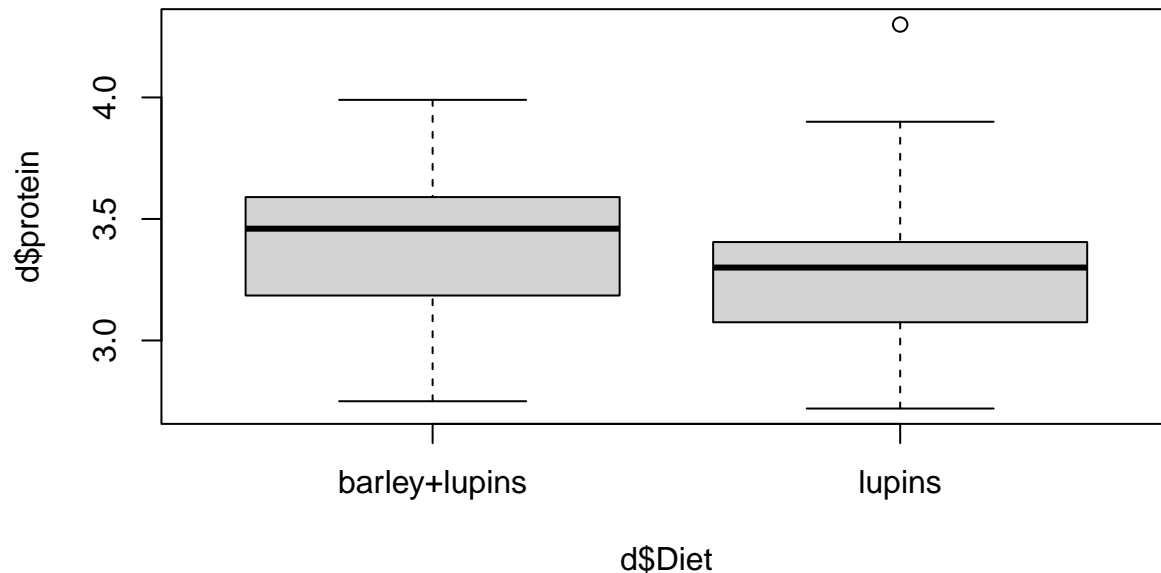
```
##  Max.   :4.300
```

Plot summarizing the data: here we can consider a stripchart because the sample size is not very big.

```
stripchart(d$protein~d$Diet,method="jitter",vertical=TRUE)
```
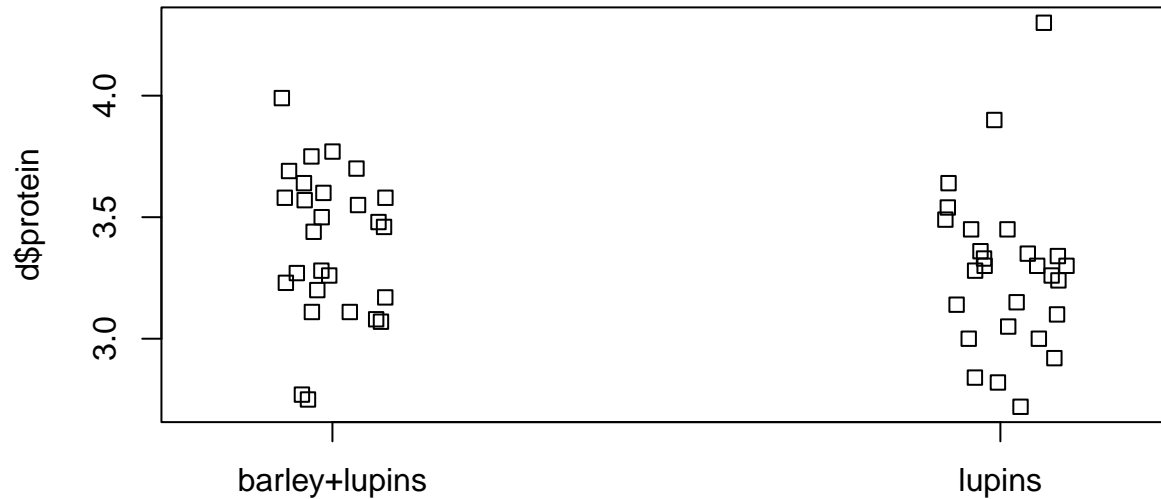

But here it is actually easier to get an idea of the data and the direction of the result with boxplots.

```
boxplot(d$protein~d$Diet)
```


Computation: first of each mean and corresponding 95% confidence interval.

```
t.test(d$protein[d$Diet=="barley+lupins"])
```

```
##
##  One Sample t-test
##
```

2

```
## data:  d$protein[d$Diet == "barley+lupins"]
## t = 58.179, df = 26, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.272728 3.512458
## sample estimates:
## mean of x
##  3.392593
```

```r
t.test(d$protein[d$Diet=="lupins"])
```

```
##
##  One Sample t-test
##
## data:  d$protein[d$Diet == "lupins"]
## t = 51.291, df = 26, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.148906 3.411835
## sample estimates:
## mean of x
##   3.28037
```

The mean protein levels are estimated to be 3.39 (95% CI=[3.27;3.51]) and 3.28 (95% CI=[3.15;3.41]) for cows fed with barley and lupins and lupins only, respectively. Note that the confidence intervals overlap in a way which enables us to conclude that the difference between the two group is not significant. This is because the estimated mean 3.28 (lupins only) is included in the 95% CI [3.27;3.51] (that of the mean for lupins and barley). But anyway, we want a p-value and a confidence interval for the difference, so it does not matter much.

```r
mean(d$protein[d$Diet=="barley+lupins"]-d$protein[d$Diet=="lupins"])
```

```
## [1] 0.1122222
```

```r
t.test(d$protein~d$Diet)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$protein by d$Diet
## t = 1.2966, df = 51.563, p-value = 0.2005
## alternative hypothesis: true difference in means between group barley+lupins and grou
## 95 percent confidence interval:
##  -0.06148816  0.28593261
## sample estimates:
## mean in group barley+lupins        mean in group lupins
```

```
##                      3.392593                        3.280370
```

We estimate the mean difference as 0.11 (95% CI=[-0.06;0.29]). The p-value is p=0.20, which is indeed larger than 5%, hence the difference is not significant. Because the confidence interval is relatively wide, it is not possible to conclude evidence against absence of effect of diet. We instead conclude that we do not know. Maybe there is no effect of the diet, maybe there is one but it is too small to lead to significant results with such a small sample size of n=27+27=54.

Visually check whether the main assumptions hold:

```r
par(mfrow=c(1,2))
qqnorm(scale(d$protein[d$Diet=="barley+lupins"]),main="barley+lupins")
abline(0,1,col="red",lty=2,lwd=3)
qqnorm(scale(d$protein[d$Diet=="lupins"]),main="lupins")
abline(0,1,col="red",lty=2,lwd=3)
```
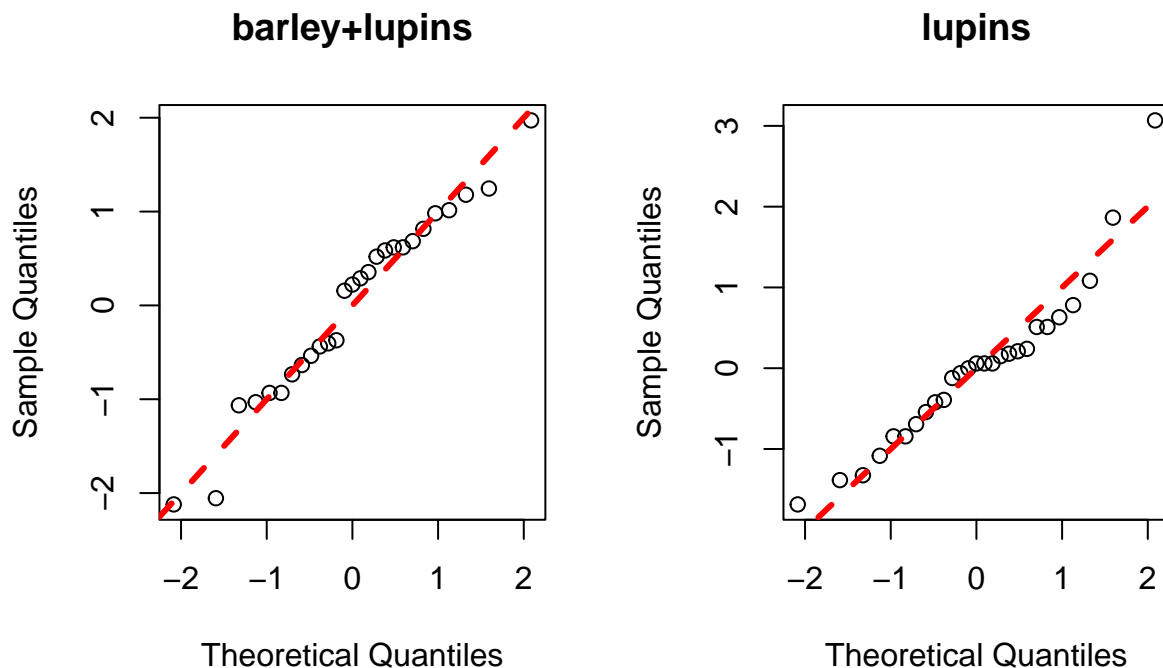

With two histograms as in the R-demo.

```r
MyCols <- c(rgb(1,0,0,alpha=0.5),rgb(0,1,0,alpha=0.5))
# Plot the first histogram
hist(d$protein[d$Diet=="barley+lupins"],
     col=MyCols[1],
     xlim=c(2,5),
     ylim=c(0,10),
     main="",
     xlab="Protein level (%)",
     axes=FALSE)
# Add the second
```

```
hist(d$protein[d$Diet=="lupins"],col=MyCols[2],add=TRUE)
# Add a legend
legend("right",
       fill=MyCols,
       legend=c("barley+lupins","lupins"),
       bty="n",
       border="white")
# Add x-axis and y-axis
axis(1)
axis(2,las=2)
```
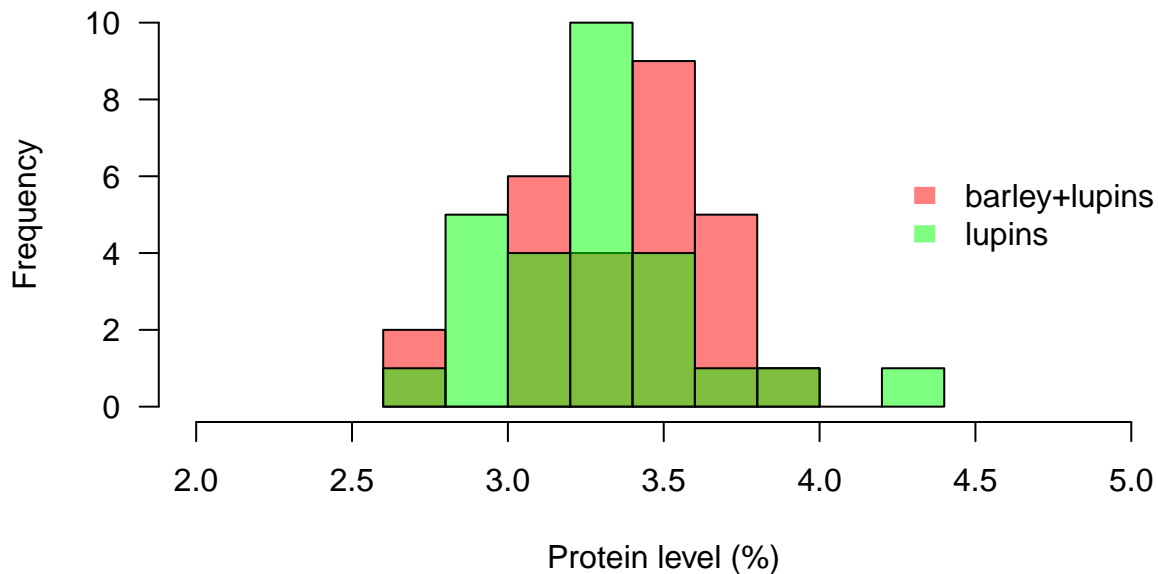

It looks approximately OK, although here it does not really matter much, because the sample size of each group is not "very small" (n=27 in each group).

Conclusion/reporting:

The mean protein level of the milk at 6 weeks after calving (expressed in %) is estimated as 3.39 (95% CI=[3.27;3.51]) and 3.28 (95% CI=[3.15;3.41]) for cows fed with barley and lupins and lupins only, respectively. The mean difference is estimated as 0.11 (95% CI=[-0.06;0.29]) and is not significantly different from 0 (p-value=0.20). Hence, we cannot conclude that the diet is associated with the mean protein level of the milk at 6 weeks after calving.

Question 2

Research hypothesis: the candidate gene NACP, coding for alpha synuclein is associated with alcohol dependence: longer allele length results in more expressed alpha synuclein mRNA.

Null and alternative hypotheses:

- H_0: the distribution of the level of expressed alpha synuclein mRNA does not depend on the allele length ("intermediate" vs "long").

- H_1: the distribution of the level of expressed alpha synuclein mRNA is different when the allele length is "intermediate" and "long".

Statistical methods: we will compare the **distribution** of the two groups using an **exact** Wilcoxon test. This test does not make any assumption regarding the distribution of the data and is exact, which means that the p-values can be trusted even with arbitrarily small sample sizes. We will also compare the two distributions visually using a dotplot.

Prepare and load the data: (similar to what is done in the R-demo of Lecture 2).

```r
library(coin)
```

```
## Loading required package: survival
```

```r
data(alpha) #
alpha <- alpha[alpha$alength!="short",]
alpha$alength <- droplevels(alpha$alength)
summary(alpha)
```

```
##          alength        elevel
##   intermediate:58   Min.   :-2.400
##   long        :15   1st Qu.: 1.750
##                     Median : 2.900
##                     Mean   : 2.487
##                     3rd Qu.: 3.500
##                     Max.   : 4.430
```

Plot summarizing the data:

```r
set.seed(123)
stripchart(alpha$elevel~alpha$alength,
           vertical=TRUE,method="jitter",
           pch=c(19,19),col=c("red","blue"),
           xlab="Allele length",ylab="Expression levels of alpha synuclein mRNA",
           axes=FALSE)
axis(1,at=1:2,c("Intermediate (n=58)","long (n=15)"))
axis(2,las=TRUE)
```

Expression levels of alpha synuclein mRNA

4
3
2
1
0
−1
−2

Intermediate (n=58)   long (n=15)

Allele length
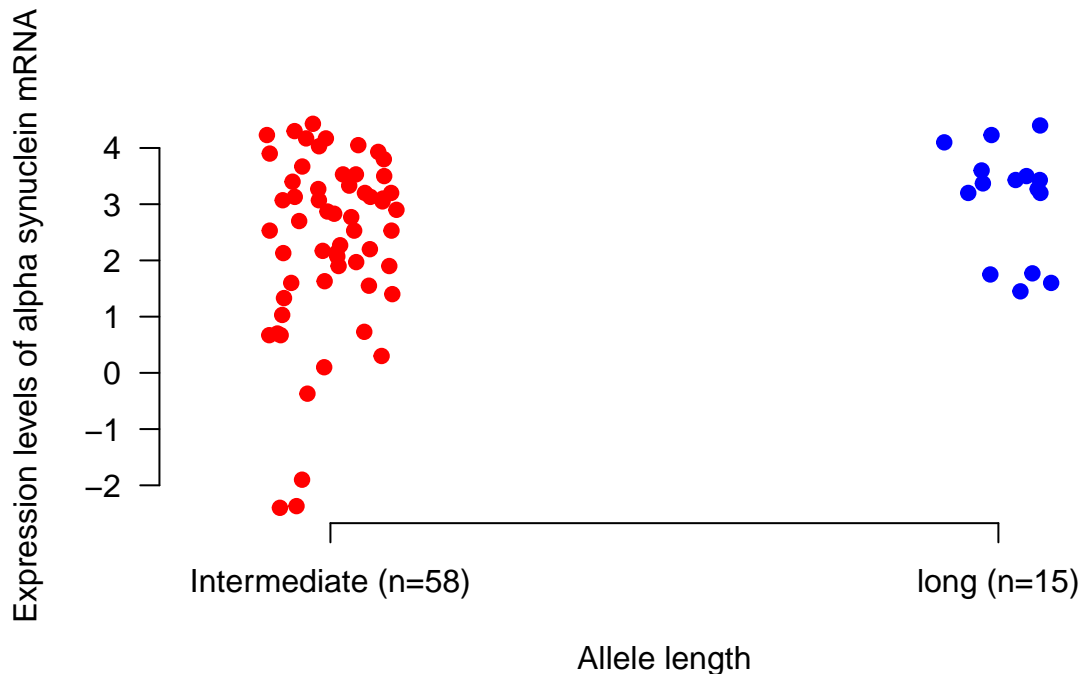
Computation: of exact Wilcoxon test.

```
wilcox_test(elevel~alength,data=alpha, distribution='exact')
```

```
##
##   Exact Wilcoxon-Mann-Whitney Test
##
## data:  elevel by alength (intermediate, long)
## Z = -1.8298, p-value = 0.06745
## alternative hypothesis: true mu is not equal to 0
```

The p-value is 0.067, not significant (at 5%).

Visually check whether the main assumptions: here there is no specific assumption to visually check. The exact Wilcoxon test does not assume normality or anything else. It only assumes that the observations are independent (within and between groups), which should be supported by the design of the study.

Conclusion/reporting: Although the dotplot suggests that there might be a difference in the distribution of the gene expression between "long" and "intermediate" allele lengths, the exact Wilcoxon test does not show a significant difference (p-value=0.067). I can be hypothesized that there is a difference, but that the sample size of the study is too small to lead to a significant result.

Question 3

Research hypothesis: the treatment has an effect on the (mean) response.

Null and alternative hypotheses: for each dose=x, with x=0.05, 0.2, 0.6 and 1 :

- H_0: the mean response is not different for dose 0 and dose x.

- H_1: the mean response is different for dose 0 and dose x.

Statistical methods: we will compare the mean response with dose 0 to that of each non zero dose using a Welch's t-test. We plan to report the estimated mean response for each dose and the differences between dose 0 and any other dose. We plan to adjust for multiple comparisons using the Bonferroni correction. We will report simultaneous 95% confidence intervals and adjusted p-values, to adjust for the fact that we perform four comparisons. This means that we will use the significance level 5%/4=1.25% instead of the usual 5% and compute 100-5%/4=98.75% confidence intervals. (Note: in Lecture 4 we will learn about an alternative, more powerful, approach).

Prepare and load the data:

```
library(DoseFinding)
```

```
## Loading required package: lattice
```

```
## Loading required package: mvtnorm
```

```
data(biom)
summary(biom)
```

```
##       dose              resp
##  Min.   :0.00    Min.   :-1.1131
##  1st Qu.:0.05    1st Qu.: 0.2595
##  Median :0.20    Median : 0.5448
##  Mean   :0.37    Mean   : 0.6990
##  3rd Qu.:0.60    3rd Qu.: 1.1397
##  Max.   :1.00    Max.   : 2.9245
```

Plot summarizing the data: here we can consider a stripchart because the sample size is not very big.

```
stripchart(biom$resp~biom$dose,
           vertical=TRUE,method="jitter",
           pch=c(15:19),col=c("red","blue","forestgreen","orange","purple"),
           xlab="Dose",ylab="Response")
```
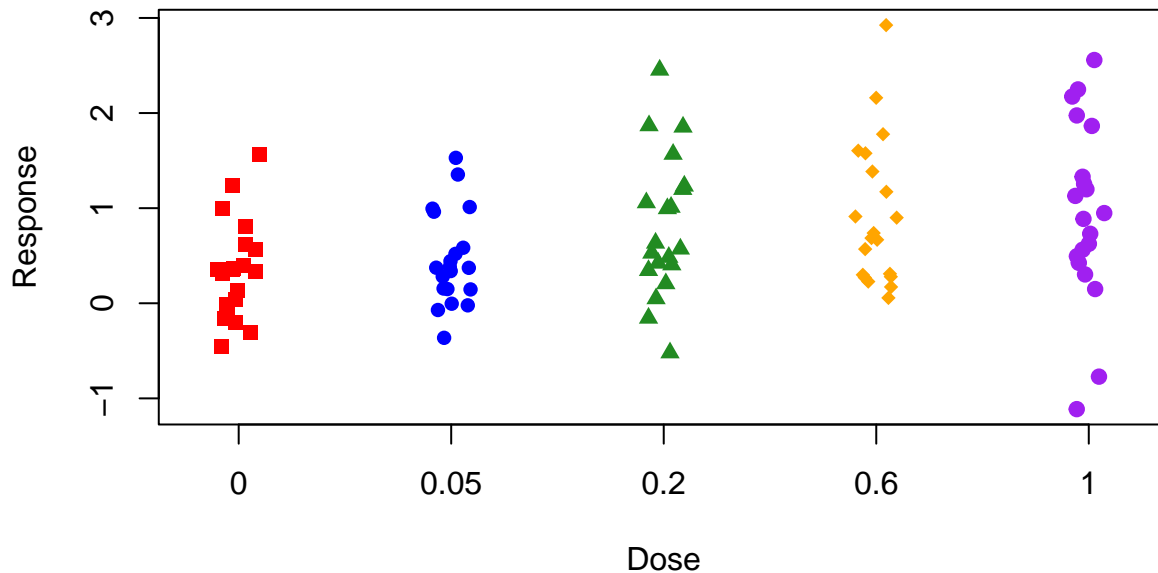
Computation: first of the mean response and corresponding 95% confidence interval for each dose. We adjust the significance level for the fact that we make four comparisons (as in R-demo).

```
K <- 4
SignLevel <- 0.05/K         # Adjusted significance level
ConfLevel <- 1 - SignLevel  # Adjusted confidence level
t.test(biom$resp[biom$dose==0],conf.level = ConfLevel)
```

```
##
##  One Sample t-test
##
## data:  biom$resp[biom$dose == 0]
## t = 2.9851, df = 19, p-value = 0.007609
## alternative hypothesis: true mean is not equal to 0
## 98.75 percent confidence interval:
##  0.02616907 0.66364173
## sample estimates:
## mean of x
## 0.3449054
```

```
t.test(biom$resp[biom$dose==0.05],conf.level = ConfLevel)
```

```
##
##  One Sample t-test
##
## data:  biom$resp[biom$dose == 0.05]
## t = 4.1663, df = 19, p-value = 0.0005242
## alternative hypothesis: true mean is not equal to 0
## 98.75 percent confidence interval:
##  0.1543252 0.7591834
```

```
## sample estimates:
## mean of x
## 0.4567543
```

```
t.test(biom$resp[biom$dose==0.2],conf.level = ConfLevel)
```

```
##
##  One Sample t-test
##
## data:  biom$resp[biom$dose == 0.2]
## t = 4.8995, df = 19, p-value = 9.955e-05
## alternative hypothesis: true mean is not equal to 0
## 98.75 percent confidence interval:
##  0.3540777 1.2665539
## sample estimates:
## mean of x
## 0.8103158
```

```
t.test(biom$resp[biom$dose==0.6],conf.level = ConfLevel)
```

```
##
##  One Sample t-test
##
## data:  biom$resp[biom$dose == 0.6]
## t = 5.4627, df = 19, p-value = 2.859e-05
## alternative hypothesis: true mean is not equal to 0
## 98.75 percent confidence interval:
##  0.4625575 1.4063163
## sample estimates:
## mean of x
## 0.9344369
```

```
t.test(biom$resp[biom$dose==1],conf.level = ConfLevel)
```

```
##
##  One Sample t-test
##
## data:  biom$resp[biom$dose == 1]
## t = 4.4782, df = 19, p-value = 0.0002576
## alternative hypothesis: true mean is not equal to 0
## 98.75 percent confidence interval:
##  0.3642971 1.5331257
## sample estimates:
## mean of x
## 0.9487114
```

Hence we get the following estimates and Bonferroni adjusted simultaneous confidence

intervals:

- Dose 0:   0.34 [0.03; 0.66]
- Dose 0.05: 0.46 [0.15; 0.76]
- Dose 0.2:   0.81 [0.35; 1.27]
- Dose 0.6:   0.93 [0.46; 1.41]
- Dose 1:   0.95 [0.36; 1.53]

We now compute the p-values for each comparison. First, we compute p-values not adjusted for multiple testing.

```
t.test(biom$resp[biom$dose==0],biom$resp[biom$dose==0.05])
```

```
##
##  Welch Two Sample t-test
##
## data:  biom$resp[biom$dose == 0] and biom$resp[biom$dose == 0.05]
## t = -0.70223, df = 37.896, p-value = 0.4868
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4343192  0.2106214
## sample estimates:
## mean of x mean of y
## 0.3449054 0.4567543
```

```
t.test(biom$resp[biom$dose==0],biom$resp[biom$dose==0.2])
```

```
##
##  Welch Two Sample t-test
##
## data:  biom$resp[biom$dose == 0] and biom$resp[biom$dose == 0.2]
## t = -2.3069, df = 33.979, p-value = 0.02729
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.87542792 -0.05539283
## sample estimates:
## mean of x mean of y
## 0.3449054 0.8103158
```

```
t.test(biom$resp[biom$dose==0],biom$resp[biom$dose==0.6])
```

```
##
##  Welch Two Sample t-test
##
## data:  biom$resp[biom$dose == 0] and biom$resp[biom$dose == 0.6]
## t = -2.8559, df = 33.35, p-value = 0.007331
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -1.0093388 -0.1697243
## sample estimates:
## mean of x mean of y
## 0.3449054 0.9344369
```

```r
t.test(biom$resp[biom$dose==0],biom$resp[biom$dose==1])
```

```
##
##  Welch Two Sample t-test
##
## data:  biom$resp[biom$dose == 0] and biom$resp[biom$dose == 1]
## t = -2.5022, df = 29.384, p-value = 0.01815
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.0970659 -0.1105461
## sample estimates:
## mean of x mean of y
## 0.3449054 0.9487114
```

They are 0.487, 0.028, 0.007 and 0.018 for comparison between Dose 0 and Dose 0.05, 0.2, 0.6 and 1, respectively. Then we apply the Bonferonni correction to these unadjusted p-values.

```r
punadj <- c(0.4868,
            0.02729,
            0.007331,
            0.01815)
p.adjust(punadj,method="bonferroni")
```

```
## [1] 1.000000 0.109160 0.029324 0.072600
```

The Bonferonni adjusted p-values are: 1, 0.109, 0.029, and 0.073. We observe a significant difference only with dose 0.6 (adjusted p-value=0.029).

We now compute the Bonferroni adjusted simultaneous confidence interval for the significant difference. That is, the 98.75% confidence intervals for the difference between Dose 0 and 0.6.

```r
mean(biom$resp[biom$dose==0])-mean(biom$resp[biom$dose==0.6])
```

```
## [1] -0.5895315
```

```r
t.test(biom$resp[biom$dose==0],biom$resp[biom$dose==0.6],conf.level = ConfLevel)
```

```
##
##  Welch Two Sample t-test
##
## data:  biom$resp[biom$dose == 0] and biom$resp[biom$dose == 0.6]
## t = -2.8559, df = 33.35, p-value = 0.007331
## alternative hypothesis: true difference in means is not equal to 0
## 98.75 percent confidence interval:
```

```
##  -1.13459103 -0.04447202
## sample estimates:
## mean of x mean of y
## 0.3449054 0.9344369
```

Visually check whether the main assumptions:

```
par(mfrow=c(2,3))
qqnorm(scale(biom$resp[biom$dose==0]),main="Dose 0")
abline(0,1,col="red",lty=2,lwd=3)
qqnorm(scale(biom$resp[biom$dose==0.05]),main="Dose 0.05")
abline(0,1,col="red",lty=2,lwd=3)
qqnorm(scale(biom$resp[biom$dose==0.2]),main="Dose 0.2")
abline(0,1,col="red",lty=2,lwd=3)
qqnorm(scale(biom$resp[biom$dose==0.6]),main="Dose 0.6")
abline(0,1,col="red",lty=2,lwd=3)
qqnorm(scale(biom$resp[biom$dose==1]),main="Dose 1")
abline(0,1,col="red",lty=2,lwd=3)
```
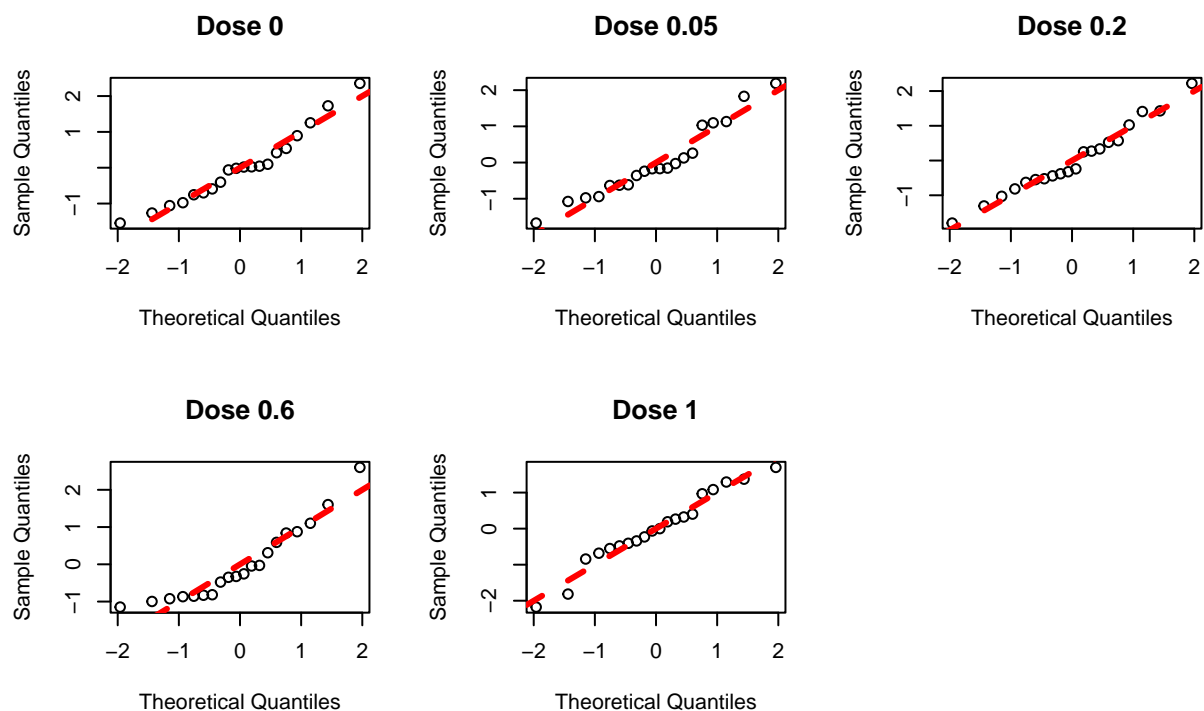

We do not observe any major deviation to what can be expected with normally distributed data. In addition, we observe n=20 responses per dose, which is still reasonable. Hence we can be quite confident with the appropriateness of the statistical method.

Conclusion/reporting: the data tend to suggest that, overall, the higher the dose the larger the mean response. However, the only significant difference that we observe (after adjusting for multiple testing) is between Dose 0 and Dose 0.6, with a difference estimated as 0.59 (98.75% CI=[1.13;0.04], adjusted p-value=0.029).

Exercise B

Question 1

```
power.t.test(delta=3,sd=2,power=0.9,sig.level=0.05)
```

```
##
##      Two-sample t test power calculation
##
##              n = 10.40147
##          delta = 3
##             sd = 2
##      sig.level = 0.05
##          power = 0.9
##    alternative = two.sided
##
## NOTE: n is number in *each* group
```

They need to include 11 mice in each arm.

Question 2

```
power.t.test(delta=3,sd=2.5,sig.level=0.05,n=11)
```

```
##
##      Two-sample t test power calculation
##
##              n = 11
##          delta = 3
##             sd = 2.5
##      sig.level = 0.05
##          power = 0.7633391
##    alternative = two.sided
##
## NOTE: n is number in *each* group
```

If they include 11 mice in each arm and the standard deviation is 2.5 instead of the anticipated value 2, then the power is only 76% (instead of the planned 90%).

```
power.t.test(delta=3,sd=3,sig.level=0.05,n=11)
```

```
##
##      Two-sample t test power calculation
##
```

```
##                 n = 11
##             delta = 3
##                sd = 3
##         sig.level = 0.05
##             power = 0.6070844
##       alternative = two.sided
##
## NOTE: n is number in *each* group
```

If they include 11 mice in each arm and the standard deviation is 3 instead of the anticipated value 2, then the power is only 61% (instead of the planned 90%).

```r
power.t.test(sd=2,sig.level=0.05,power=0.75,n=11)
```

```
##
##         Two-sample t test power calculation
##
##                 n = 11
##             delta = 2.36173
##                sd = 2
##         sig.level = 0.05
##             power = 0.75
##       alternative = two.sided
##
## NOTE: n is number in *each* group
```

If they include 11 mice in each arm and assuming that the standard deviation is 2, then they can hope to show a difference as small as 2.36, with a power of 75%.