# Exercise 4 - solution

## Paul Blanche

## Exercise A

### Question 1

We first load the data.

```r
library(DoseFinding)
```

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

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

```r
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
```

From the above summary, we can see that R thinks of the variable dose as a numeric variable. We therefore create a new variable that considers the variabe dose as a factor variable (i.e. a categorical variable). We then print the frequency of each level.

```r
biom$dosefact <- factor(biom$dose)
table(biom$dosefact)
```

```
##
##    0 0.05  0.2  0.6    1
##   20   20   20   20   20
```

We can now fit an ANOVA model and compute the estimated differences, 95% limits and p-values adjusted for multiple comparisons (four doses are compared to dose 0).

```
library(multcomp) # required package
```

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

```
library(sandwich) # required package
fitlm <- lm(resp~dosefact, data=biom) # fit a linear model (here a 1-way ANOVA)
glht.Dunnett <- glht(fitlm,
                     mcp(dosefact="Dunnett"), # for "many-to-one" comparisons
                     vcov=vcovHC) # Do not assume equal variances in all groups.
summary(glht.Dunnett) # print estimated differences and adjusted p-values
```

```
##
##     Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = resp ~ dosefact, data = biom)
##
## Linear Hypotheses:
##                Estimate Std. Error t value Pr(>|t|)
## 0.05 - 0 == 0    0.1118     0.1634   0.684   0.9184
## 0.2 - 0 == 0     0.4654     0.2070   2.248   0.0951 .
## 0.6 - 0 == 0     0.5895     0.2118   2.784   0.0244 *
## 1 - 0 == 0       0.6038     0.2476   2.439   0.0603 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
confint(glht.Dunnett) # print 95% limits
```

```
##
##     Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
```

2

```
## Fit: lm(formula = resp ~ dosefact, data = biom)
##
## Quantile = 2.5137
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                Estimate lwr      upr
## 0.05 - 0 == 0  0.11185 -0.29894  0.52263
## 0.2 - 0 == 0   0.46541 -0.05492  0.98574
## 0.6 - 0 == 0   0.58953  0.05715  1.12191
## 1 - 0 == 0     0.60381 -0.01855  1.22616
```

**Note:** by default, the reference level (here dose 0) is the level which is compared to all others. If you want to compare a different level to all others, you can do it by first changing the reference level using the `relevel` function.

In exercise 2, using the Bonferroni correction we saw that only the comparison between dose 0 and dose 0.6 led to a significant result. This is similar here. However, the new results provide lower p-values and narrower confidence intervals. For the comparison of doses 0 and 0.6:

- Bonferroni lead to: p-value=0.029, confidence limits=[0.045;1.135]
- min-P approach lead to: p-value=0.024, confidence limits=[0.057;1.122]

However, the differences are not very big here and do not change the overall conclusion.

## Question 2

```
glht.Tukey <- glht(fitlm, mcp(dosefact="Tukey"), vcov=vcovHC)
summary(glht.Tukey)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = resp ~ dosefact, data = biom)
##
## Linear Hypotheses:
##                Estimate Std. Error t value Pr(>|t|)
## 0.05 - 0 == 0   0.11185    0.16342   0.684   0.9579
## 0.2 - 0 == 0    0.46541    0.20699   2.248   0.1658
## 0.6 - 0 == 0    0.58953    0.21179   2.784   0.0476 *
```

```
## 1 - 0 == 0          0.60381    0.24758   2.439   0.1100
## 0.2 - 0.05 == 0     0.35356    0.20358   1.737   0.4081
## 0.6 - 0.05 == 0     0.47768    0.20845   2.292   0.1516
## 1 - 0.05 == 0       0.49196    0.24474   2.010   0.2619
## 0.6 - 0.2 == 0      0.12412    0.24412   0.508   0.9858
## 1 - 0.2 == 0        0.13840    0.27575   0.502   0.9865
## 1 - 0.6 == 0        0.01427    0.27937   0.051   1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

We can see that the p-values for the comparisons also considered in the previous question are bigger (e.g. dose 0 vs dose 0.6, p-value is now 0.048, whereas it was 0.029). This makes perfect sense. When we consider more comparisons we need to adjust the p-value more, i.e. we need to inflate the p-value more to compensate for the additional risk of false positive results. Although here the comparison between dose 0 and 0.6 is still significant, this is now "borderline".

**Take home message**: do not make more comparisons than necessary! (For the main analysis). Indeed, either you will rigorously adjust for multiple testing but then you will have lower chances of true positive findings or you will not rigorously adjust for multiple testing but you will have a higher risk of false positive findings.

```
confint(glht.Tukey)
```

```
##
##    Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = resp ~ dosefact, data = biom)
##
## Quantile = 2.7656
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                  Estimate  lwr        upr
## 0.05 - 0 == 0    0.111849 -0.340096   0.563793
## 0.2 - 0 == 0     0.465410 -0.107051   1.037872
## 0.6 - 0 == 0     0.589532  0.003809   1.175254
## 1 - 0 == 0       0.603806 -0.080909   1.288522
## 0.2 - 0.05 == 0  0.353561 -0.209461   0.916584
## 0.6 - 0.05 == 0  0.477683 -0.098819   1.054184
## 1 - 0.05 == 0    0.491957 -0.184887   1.168801
```

```
## 0.6 - 0.2 == 0    0.124121 -0.551017  0.799259
## 1 - 0.2 == 0      0.138396 -0.624216  0.901007
## 1 - 0.6 == 0      0.014274 -0.758341  0.786890
```

For similar reasons, we observe larger confidence limits.

For comparison and pedagogical purpose, we now compute the unadjusted p-value for the comparison of doses 0 and 0.6.

```
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
```

The (unadjusted) p-value is 0.007331. When we adjust using Bonferroni, we multiply it by 10 (because we make 10 comparisons here). This leads to p-value=0.073, which is not significant (whereas this comparison was significant according to the min-P approach).

This is just an example showing that using the "modern and efficient" min-P approach instead of the simpler "good old" Bonferroni correction can matter.

**Take home message**: a sub-optimal choice of the statistical method can lower your chance of showing positive results, while it should **not** increase your risk of false positive results.

# Exercise B

# Part 1

## Question 1

We first load the data.

```
load(url("http://paulblanche.com/files/SCD.rda"))
d <- SCD # shorter name, just for convenience
head(d)
```

```
##   age sex SCD Pdias Psys height weight pulse  HCT Creat  Hb
```

```
## 1  26   2   1    62  120     164    57.0    68 21.5    37 7.7
## 2  33   2   1    60  115     165    58.0    64 21.6    55 7.8
## 3  43   2   1    66  117     152    57.2    67 29.3    68 9.2
## 4  23   1   1    67  129     177    61.0    62 25.0    66 8.9
## 5  37   1   1    47  106     167    50.0    65 23.1    66 8.0
## 6  46   2   1    65  134     163    68.0    67 21.0    69 7.7
```

Then we create the variable MAP as in day 1.

```
d$MAP <- d$Pdias + (1/3)*(d$Psys-d$Pdias)
head(d)
```

```
##   age sex SCD Pdias Psys height weight pulse  HCT Creat  Hb      MAP
## 1  26   2   1    62  120     164    57.0    68 21.5    37 7.7 81.33333
## 2  33   2   1    60  115     165    58.0    64 21.6    55 7.8 78.33333
## 3  43   2   1    66  117     152    57.2    67 29.3    68 9.2 83.00000
## 4  23   1   1    67  129     177    61.0    62 25.0    66 8.9 87.66667
## 5  37   1   1    47  106     167    50.0    65 23.1    66 8.0 66.66667
## 6  46   2   1    65  134     163    68.0    67 21.0    69 7.7 88.00000
```

Then we create the BMI variable and its categorized version as in day 1. Finally, we print the counts.

```
d$BMI <- d$weight/(d$height/100)^2
d$BMIgroup <- cut(d$BMI,
                  breaks=c(0,18.5,25,30,Inf),
                  include.lowest=TRUE,right=FALSE)
table(d$BMIgroup)
```

```
##
##  [0,18.5) [18.5,25)   [25,30)  [30,Inf]
##        32       112        23         9
```

## Question 2

We first fit a "usual" ANOVA model that assume the same variance in all groups.

```
lmfit <- lm(MAP~BMIgroup,data=d)
summary(lmfit)
```

```
##
## Call:
## lm(formula = MAP ~ BMIgroup, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.3333  -5.3910  -0.4077   4.9312  29.2589
```

```
## 
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          77.385      1.552  49.853  < 2e-16 ***
## BMIgroup[18.5,25)     6.356      1.760   3.611    4e-04 ***
## BMIgroup[25,30)      18.281      2.400   7.616 1.67e-12 ***
## BMIgroup[30,Inf]     19.874      3.313   5.998 1.14e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 8.781 on 172 degrees of freedom
## Multiple R-squared:  0.3114, Adjusted R-squared:  0.2994
## F-statistic: 25.93 on 3 and 172 DF,  p-value: 6.896e-14
```

From the output, we can see that the estimated mean MAP is:

- 77.39 in the reference group, here BMI < 18.5.
- 77.39 + 6.36 = 83.74 in the group of BMI in [18.5,25)
- 77.39 + 18.28 = 95.67 in the group of BMI in [25,30)
- 77.39 + 19.87 = 97.26 in the group of BMI > 30

**Remember:** here the intercept is the mean in the reference group and the other estimates should be interpreted as estimated differences to the reference group.

In the output, we can also read "Residual standard error: 8.781". This means that the estimate of the standard **deviation** of the "error term" is 8.78. For interpreting this value, we can for example say the following. For any subject picked up at random in a population similar to that from which our sample was drawn, 95 times out of 100, his/her MAP will be observed not further away than 2 times 8.78 (i.e. 17.56) from the mean values estimated for his/her BMI group. This interpretation relies on the normal distribution of the "error term". But even if this assumption does not hold, this value 8.7 still gives us some information about the spread of MAP observations around the mean. For instance, we can still make the same interpretation but replacing "95 times out of 100" by "at least 75 times out of 100" in the above sentence.

We now compute the (sample) mean of MAP in each BMI group.

```
tapply(d$MAP,d$BMIgroup,mean)
```

```
##  [0,18.5) [18.5,25)   [25,30)  [30,Inf]
##  77.38542  83.74107  95.66667  97.25926
```

The (sample) mean of MAP in each BMI group match with the model estimates. This makes sense as they can both be used to estimate the population mean MAP in each group.

We now compute the (sample) standard deviation of MAP in each BMI group.

```
tapply(d$MAP,d$BMIgroup,sd)
```

```
##  [0,18.5) [18.5,25)   [25,30)  [30,Inf]
```

```
##  7.869504  8.200529 12.076356  9.148129
```

The (sample) standard deviations in each BMI group are relatively close to 8.78, the estimate of the standard **deviation** of the "error term", except maybe for the group of BMI in [25;30]. However, we should keep in mind that we do not observe a lot of observations for all BMI groups (n=32, 112, 23 and 9, respectively, see question 1). Hence sampling random variation might explain the difference from one BMI group to another. Finally, note that the model estimate 8.78 can be seen as a "weighted" average of standard deviations in each BMI group, with heavier weights for large groups (8.78 is close to the value of the largest group and within the range of all values).

## Question 3

We first produce the "default" model checking plots.

```
par(mfrow=c(1,2))
plot(lmfit,ask=FALSE,which=1:2)
```
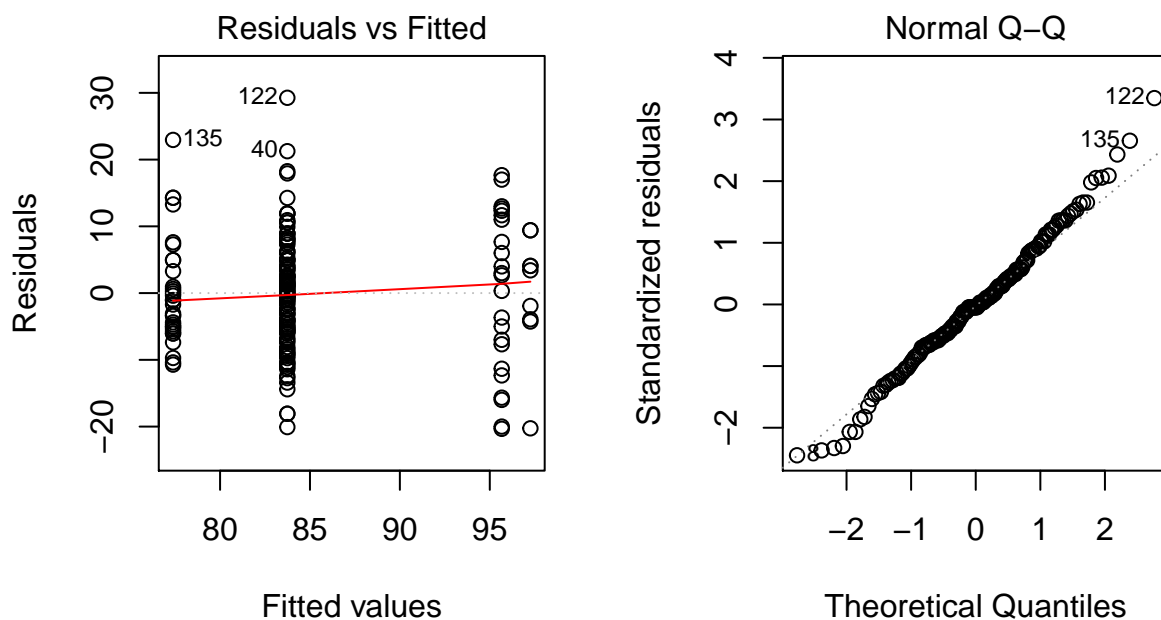


```
par(mfrow=c(1,1))
```

**Right plot**. Nothing seems very bad and anyway the normality assumption is not very important when we have large sample sizes and when we are not interested in prediction intervals, but only on estimating means and computing corresponding confidence intervals and p-values.

**Left plot**. This plot matters more, and is about checking homogeneity (equal variances accross groups). A first casual look might suggest that the spread of the residual is not as large for the largest fitted values, i.e. the largest BMI group. However, this is not really true, because they are actually few observed residuals close to 0. This "dotplot" without

"jitter" should therefore be interpreted very carefully. A "Wally plot" might help and we now produce one.

```
library(MESS)
permsr <- function(n) {sample(x=residuals(lmfit),size=n,replace=FALSE)}
permsrplot <- function(x, y, ...) {
    plot(x, y,ylab="Residuals",xlab="Fitted values",  ...) ;
    abline(h=0,lty=3,col="grey")
    lines(lowess(x, y), lty=1,col="red")}
    wallyplot(predict(lmfit),
        residuals(lmfit),
        FUN=permsrplot,simulateFunction=permsr,hide=FALSE,col="blue")
```
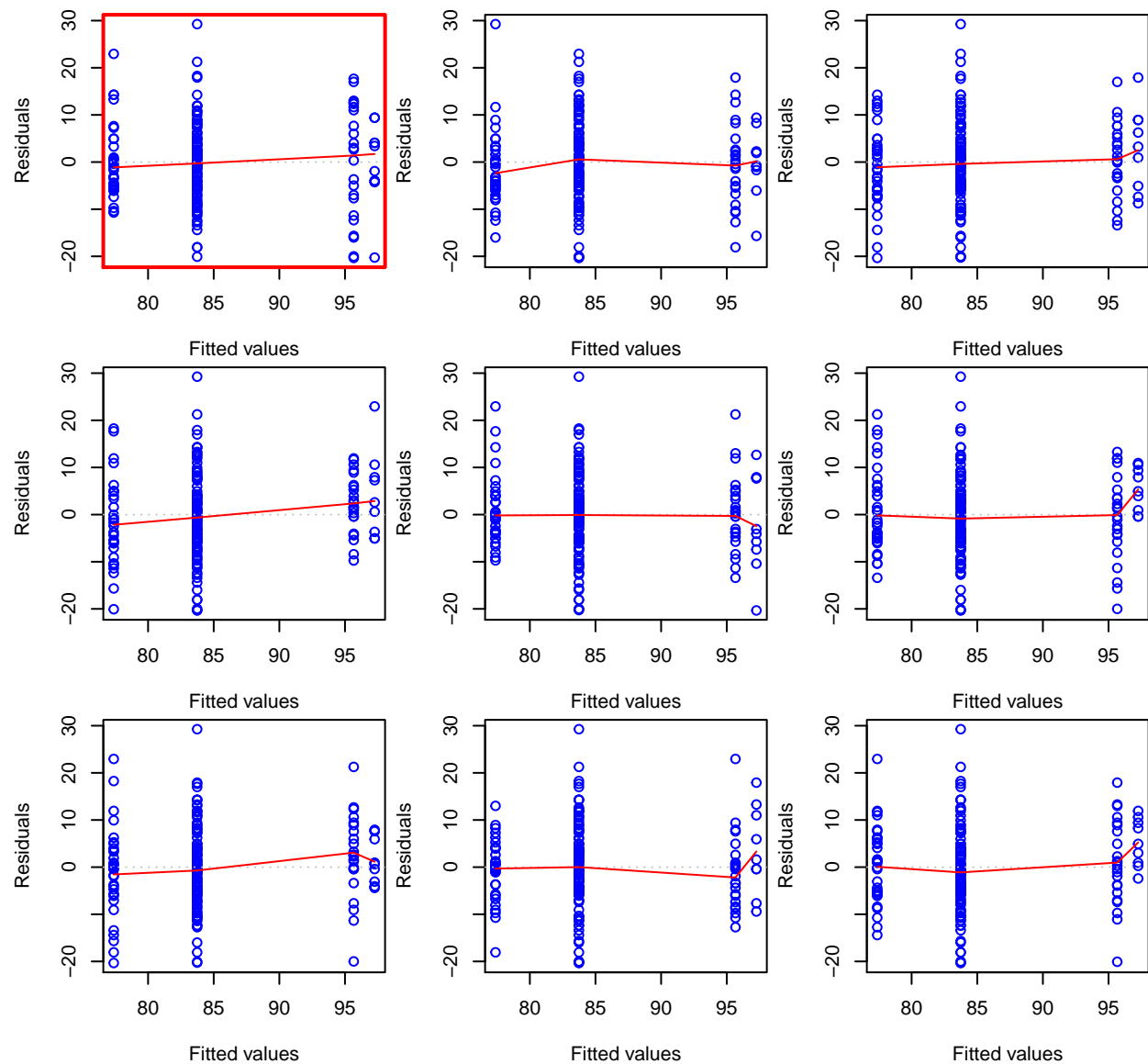


The main conclusion is that the plot corresponding to our observed data does not differ in any important/peculiar way from what can be expected due to random sampling variation.

9

From this plot, it becomes evident that e.g. what we observe for the largest BMI group might be due to random sampling variation only. That seems to be the case also for the second largest BMI group. Hence we can reasonably conclude that there is nothing specific that seem to contradict our main model assumtpion: homogeneity (i.e. equal variance of MAP accross BMI groups).

## Question 4

From the output in question 2, we can read "F-statistic: 25.93 on 3 and 172 DF, p-value: 6.896e-14". this means that the p-value of the F-test is p-value < 0.0001. Hence we conclude to a significant association between BMI and MAP.

We could have preferred the F-test to the min-P approach here because of the following. The F-test approach is a "good-old" method which is commonly used and considered by many as a simpler approach to the min-P approach. As statistical guidelines of many medical journals (e.g. European Heart Journal) suggest to use methods 'as simple as possible, but as sophisticated as needed', this can favor the F-test here. Indeed, although we can reasonably argue that the min-P approach is often more appropriate when the aim is to report the comparisons between all or some of the groups, this was not the aim here. Here we just wanted to investigate an "overall association" and nothing else. Hence the min-P approach is somehow less attractive.

## Question 5

Some people sometimes recommend to use BMI as a continuous variable instead of categorizing it, for the statistical analysis. Some go as far as saying that this is always a "better and more powerful" approach. This statement is, however, seen by many as being a bit naive. One main disadvantage of using BMI as a continuous variable is that it is then not so easy to fit a "good" model. Assuming linearity does not seem realistic, neither from a clinical point of view nor from what the data suggest. Hence more complicated modeling approaches would be needed (e.g. using splines), which are beyond the scope of this course. These approaches all have pros and cons...

# Part 2

## Question 6

We first transform the sex variable into a factor variable.

```
d$sex <- factor(d$sex,levels=c(1,2),labels=c("male","female"))
table(d$sex)
```

```
##
##   male female
##     82     94
```

## Question 7

```
lm2way <- lm(MAP~BMIgroup+sex,data=d)
summary(lm2way)
```

```
##
## Call:
## lm(formula = MAP ~ BMIgroup + sex, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.1600  -4.6928   0.0382   5.1489  27.7048
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         78.886      1.669  47.267  < 2e-16 ***
## BMIgroup[18.5,25)    6.409      1.739   3.685 0.000306 ***
## BMIgroup[25,30)     18.607      2.376   7.832 4.85e-13 ***
## BMIgroup[30,Inf]    20.374      3.281   6.210 3.89e-09 ***
## sexfemale           -3.001      1.316  -2.280 0.023850 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.676 on 171 degrees of freedom
## Multiple R-squared:  0.3317, Adjusted R-squared:  0.3161
## F-statistic: 21.22 on 4 and 171 DF,  p-value: 3.176e-14
```

We estimate that, for any BMI group, the average difference in MAP between a woman and a man **that belong to the same BMI group** is -3.00 mmHg. In other words, men have in average a MAP 3 mmHg higher than women of the same BMI group.

According to a model assumption, this average difference between men and women is the same whatever the BMI group. This assumption is assumed to be supported by subject-matter knowledge and will be checked.

It was a good idea to transform the variable sex into a factor variable, to avoid confusion and mistakes. Here is what would have happened otherwise.

```
d$sexNumeric <- as.numeric(d$sex=="female") + 1 # recreate original variable
table(d$sexNumeric)
```

```
##
```

```
## 1 2
## 82 94
```

```
lm2wayNum <- lm(MAP~BMIgroup+sexNumeric,data=d)
summary(lm2wayNum)
```

```
##
## Call:
## lm(formula = MAP ~ BMIgroup + sexNumeric, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.1600  -4.6928   0.0382   5.1489  27.7048
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         81.887      2.500  32.753  < 2e-16 ***
## BMIgroup[18.5,25)    6.409      1.739   3.685 0.000306 ***
## BMIgroup[25,30)     18.607      2.376   7.832 4.85e-13 ***
## BMIgroup[30,Inf]    20.374      3.281   6.210 3.89e-09 ***
## sexNumeric          -3.001      1.316  -2.280 0.023850 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.676 on 171 degrees of freedom
## Multiple R-squared:  0.3317, Adjusted R-squared:  0.3161
## F-statistic: 21.22 on 4 and 171 DF,  p-value: 3.176e-14
```

First of all, with this output it is less clear what the estimate -3.001 at the line "sexNumeric" means. It is the difference in mean for a one unit increase of the variable sex, when other variables remain similar. But, first, we are lucky that sex was coded 1 vs 2, not e.g. 1 vs 3, otherwise we would have obtained a different estimate and should have multiplied it by 2 for a correct interpretation. Second, it is less clear whether -3.001 is the difference estimated for men vs women or for women vs men, i.e. the sign is less easy to interpret because we need to remember what sex=1 and sex=2 mean. Third, the estimate of the intercept does not have a nice interpretation anymore!

**Take home message**: it is much safer (if not mandatory) to use factor variables in the **lm** function for variables coding categorical variables.

From the output, we can read that the estimated standard deviation of the "error term" is 8.676. This is smaller that 8.78 obtained in Part 1, as expected, because part of the variation of MAP is now explained by sex. The interpretation in terms of spread around the estimated means is the same as in question 2 of Part 1.

## Question 8

We create a new variable "group", to simultaneously indicate both the sex and the BMI group of the subject corresponding to each observation.

```
d$group <- interaction(d$sex,d$BMIgroup)
table(d$group)
```

```
##
##    male.[0,18.5)  female.[0,18.5)   male.[18.5,25) female.[18.5,25)
##              16               16               54               58
##    male.[25,30)  female.[25,30)    male.[30,Inf]  female.[30,Inf]
##               9               14                3                6
```

We now produce boxplots of MAP according to the BMI and sex groups.

```
boxplot(d$MAP~d$group,col=rep(c("blue","pink"),4))
```
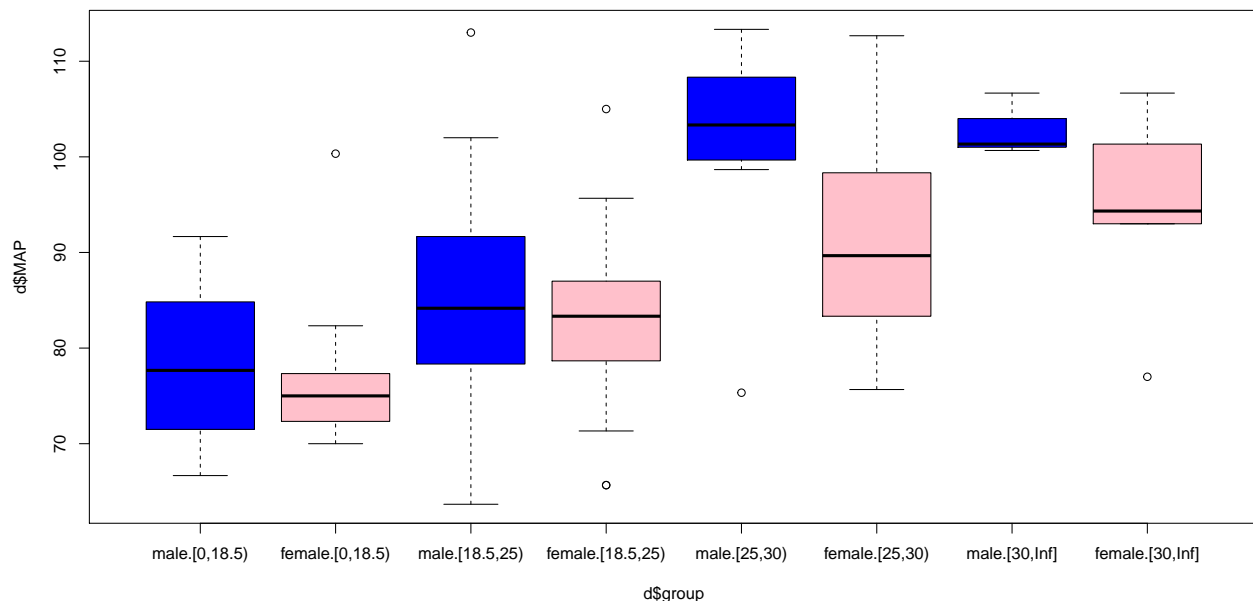


Overall, the boxplots match with the main results of the ANOVA model. This is because, overall, it seems that:

- whatever the sex, larger BMIs seem associated with higher mean MAP. Both pink and blue boxplots are located higher and higher with increasing BMI. This matches with increasing differences estimated by the model (6.4, 18.6 and 20.4)
- whatever the BMI group, men have in average a MAP higher than women of the same BMI group. For each BMI group, the blue boxplot is located higher than the pink. This matches with the differences estimated by the model (-3.0).

The main assumptions of our two-way ANOVA model are:

- homogeneity (i.e. equal variances in all subgroups).
- no interaction between sex and BMI group.

The boxplots do not seem to contradict them, at least not "strongly". Homogeneity does not seem to be a violated because the spread of the observations does not seem very different from one BMI and gender group to another, when looking at group of reasonable sizes (which are not too much impacted by small sample random variation). The assumption of no interaction does not seem to be strongly violated either, because the differences observed between the pink and blue boxplots are roughly the same for all BMI groups (again when keeping the sample size and associated random variation in mind).

## Question 9

We first produce the "default" model checking plots.

```
par(mfrow=c(1,2))
plot(lm2way,ask=FALSE,which=1:2)
```
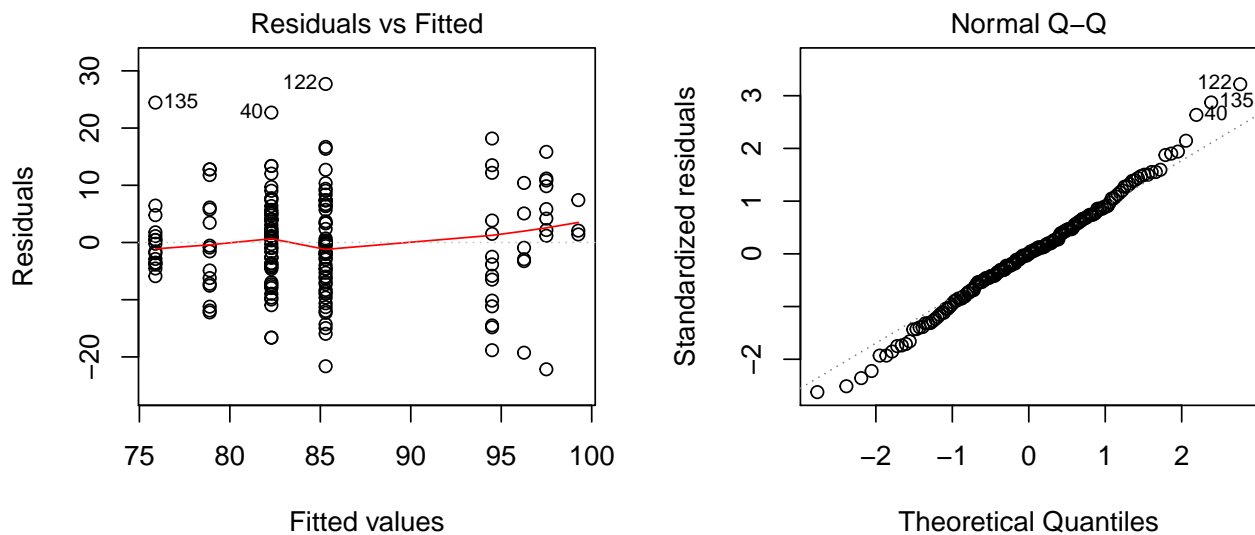


```
par(mfrow=c(1,1))
```

Again, we produce a Wally plot because we find it helpful for a more thorough understanding.

```
permsr <- function(n) {sample(x=residuals(lm2way),size=n,replace=FALSE)}
permsrplot <- function(x, y, ...) {
    plot(x, y,ylab="Residuals",xlab="Fitted values",  ...) ;
    abline(h=0,lty=3,col="grey")
    lines(lowess(x, y), lty=1,col="red")}
wallyplot(predict(lm2way),residuals(lm2way),
          FUN=permsrplot,
          simulateFunction=permsr,
          hide=FALSE,col="blue")
```

14
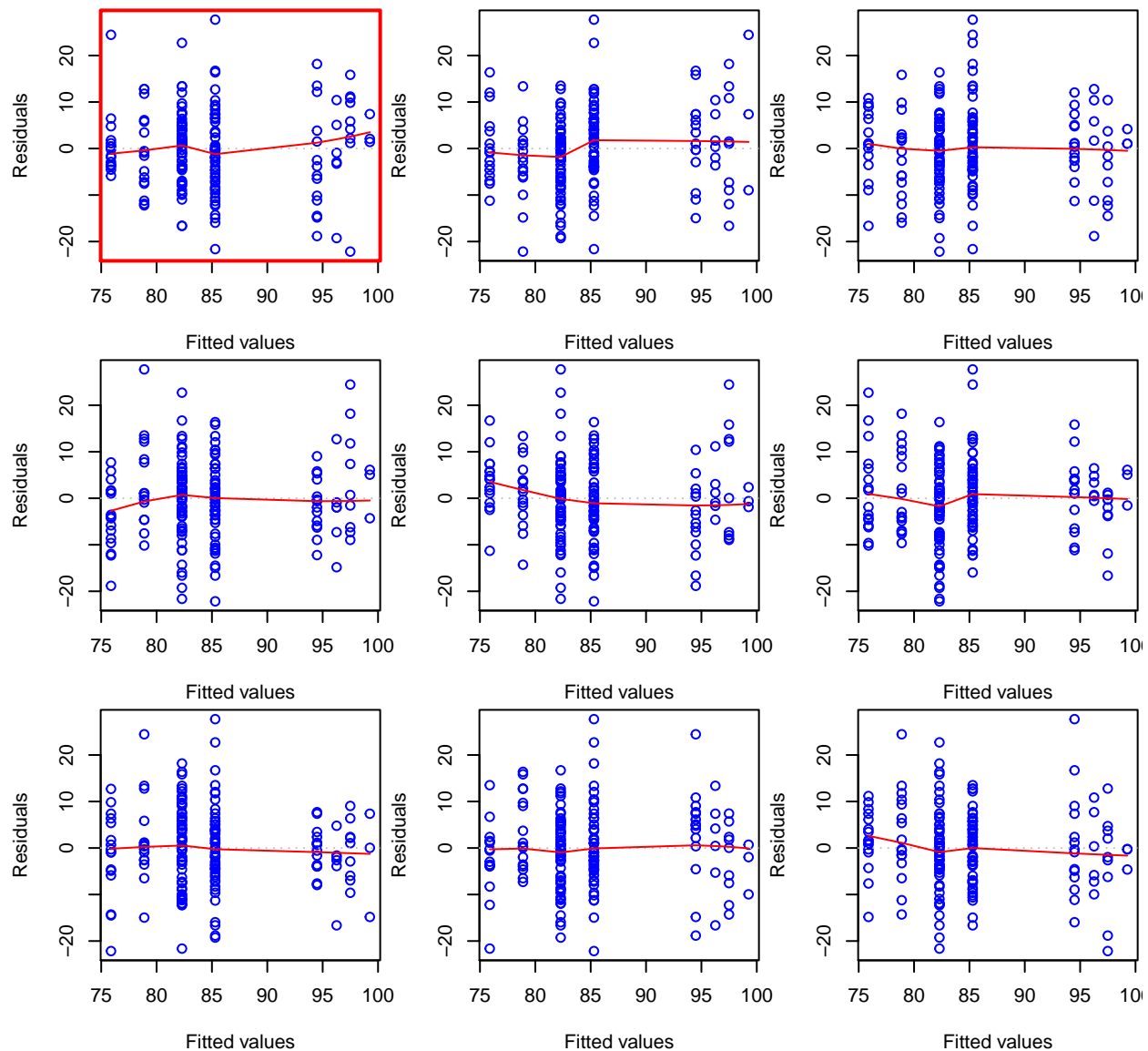
Again, here we do not see anything seriously worrying.

## Question 10

From the output seen at Question 7, we can see that we estimate a significant mean MAP between men and women of the same BMI group (p-value=0.024).

We could also compute the corresponding 95% confidence interval (-5.60; -0.40).

```
confint(lm2way)
```

```
##                         2.5 %      97.5 %
## (Intercept)         75.591555 82.1802771
## BMIgroup[18.5,25)    2.976199  9.8422891
```

```
## BMIgroup[25,30)    13.917434 23.2974566
## BMIgroup[30,Inf]   13.897941 26.8500772
## sexfemale          -5.599273 -0.4027248
```

We can produce the p-value of the F-test, which is here equal to the above p-value because there is only one parameter (because only two levels exist for sex).

```
anova(lm(MAP~BMIgroup+sex,data=d), lm(MAP~BMIgroup,data=d))
```

```
## Analysis of Variance Table
##
## Model 1: MAP ~ BMIgroup + sex
## Model 2: MAP ~ BMIgroup
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    171 12871
## 2    172 13262 -1   -391.24 5.1979 0.02385 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can produce the p-value of the min-P test, which is here also equal to the above p-value because there is only one parameter (because only two levels exist for sex) and hence we perform only one hypothesis test. Hence the multiple testing correction vanishes.

```
TwoWay.mc <- glht(lm2way, linfct = mcp(sex = "Tukey"))
summary(TwoWay.mc)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = MAP ~ BMIgroup + sex, data = d)
##
## Linear Hypotheses:
##                   Estimate Std. Error t value Pr(>|t|)
## female - male == 0   -3.001      1.316   -2.28   0.0238 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

## Question 11

Instead of using the two-way ANOVA model, we now use a simpler t-test to compare the mean MAP of men and women (just for academic purpose).

```
t.test(d$MAP~d$sex)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$MAP by d$sex
## t = 1.2361, df = 156.81, p-value = 0.2183
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.185581  5.151504
## sample estimates:
##   mean in group male mean in group female
##             85.89431             83.91135
```

The difference is not significant, when using a t-test. This is the first difference with the results of two-way ANOVA approach. Another difference is the estimate of the difference. It is -3.001 with the two-way ANOVA approach, but only 83.91135-85.89431= -1.98, when comparing sample means (as in the t-test). Although overall the result is in the same direction (MAP higher for men than women), the effect size is smaller. The difference can be better understood by looking at the pink and blue boxplots of question 8. The two-way ANOVA model compares men to women **within BMI groups**. To some extent, the estimated sex difference is an "average" of the difference in each BMI groups. However, when we compare men to women overall, what we observe in the the group "BMI in [18.5,25)", which contains 112 observations out of n=176 (i.e. 64% of the data), has a very big influence on the results. However, within this group the difference between men and women is not very large (-1.63 only).

# Appendix

## ggplot Q8

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
d %>%
    ggplot(., aes(y = MAP, x = group, fill = sex)) +
    stat_boxplot(geom ='errorbar') +
    geom_boxplot(fill = rep(c("blue", "pink"), 4)) +
    theme_classic() +
    theme(legend.title = element_blank())
```