STA303/1002 - Week 4 R Markdown

Jan 28-Feb 1, 2019

Case Study 2: The Data

Get the data (from R library):

```
#load Sleuth3 R data library; see case1302
library(Sleuth3)
#Pygmalion data
pyg = case1302
attach(pyg)
head(pyg)
```

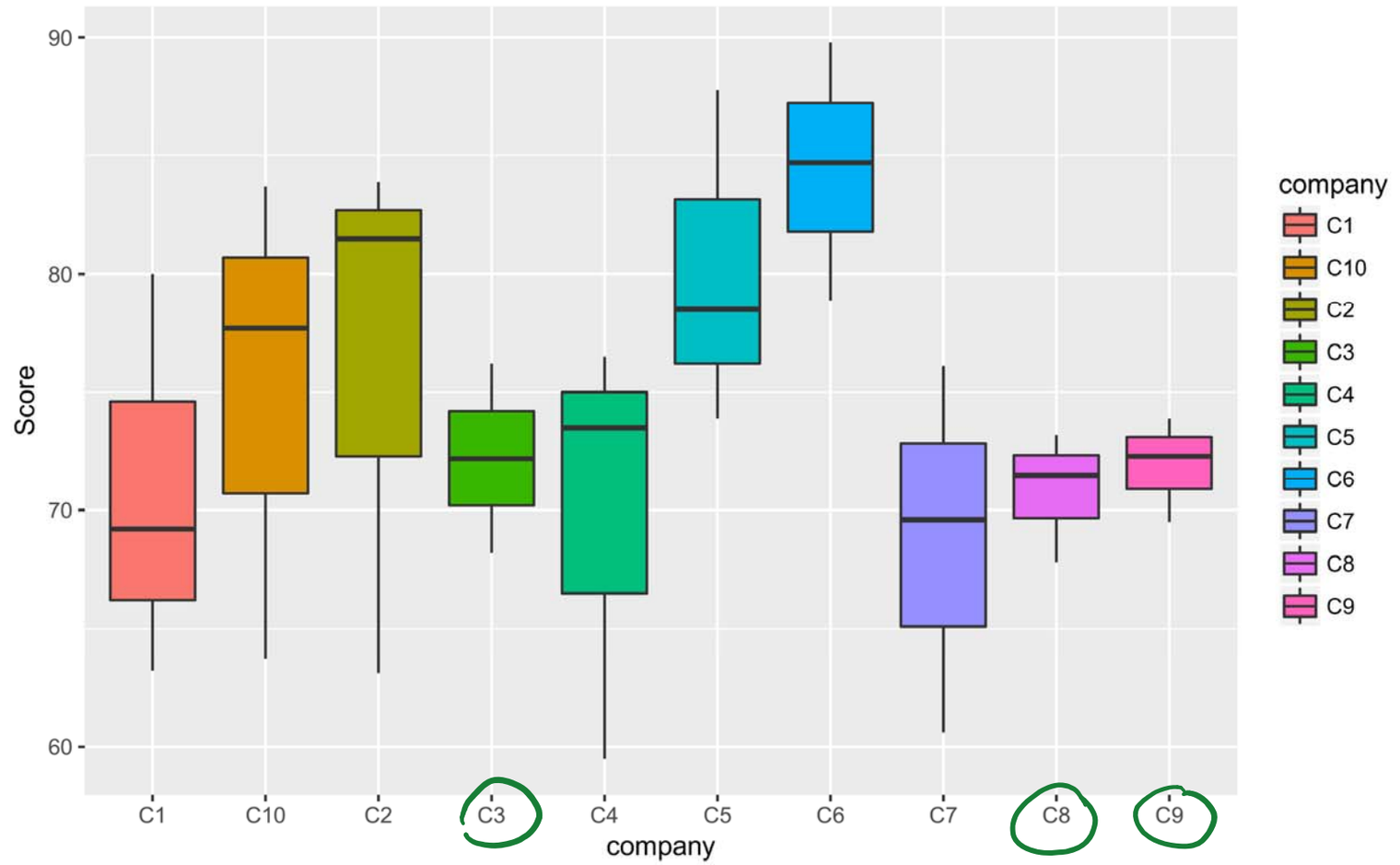
In R:
tail(pyg)
str(pyg)
dim(pyg)

```
##   Company     Treat Score
## 1      C1 Pygmalion  80.0
## 2      C1   Control  63.2
## 3      C1   Control  69.2
## 4      C2 Pygmalion  83.9
## 5      C2   Control  63.1
## 6      C2   Control  81.5
```

```
company=as.factor(Company)
treat=as.factor(Treat)
```

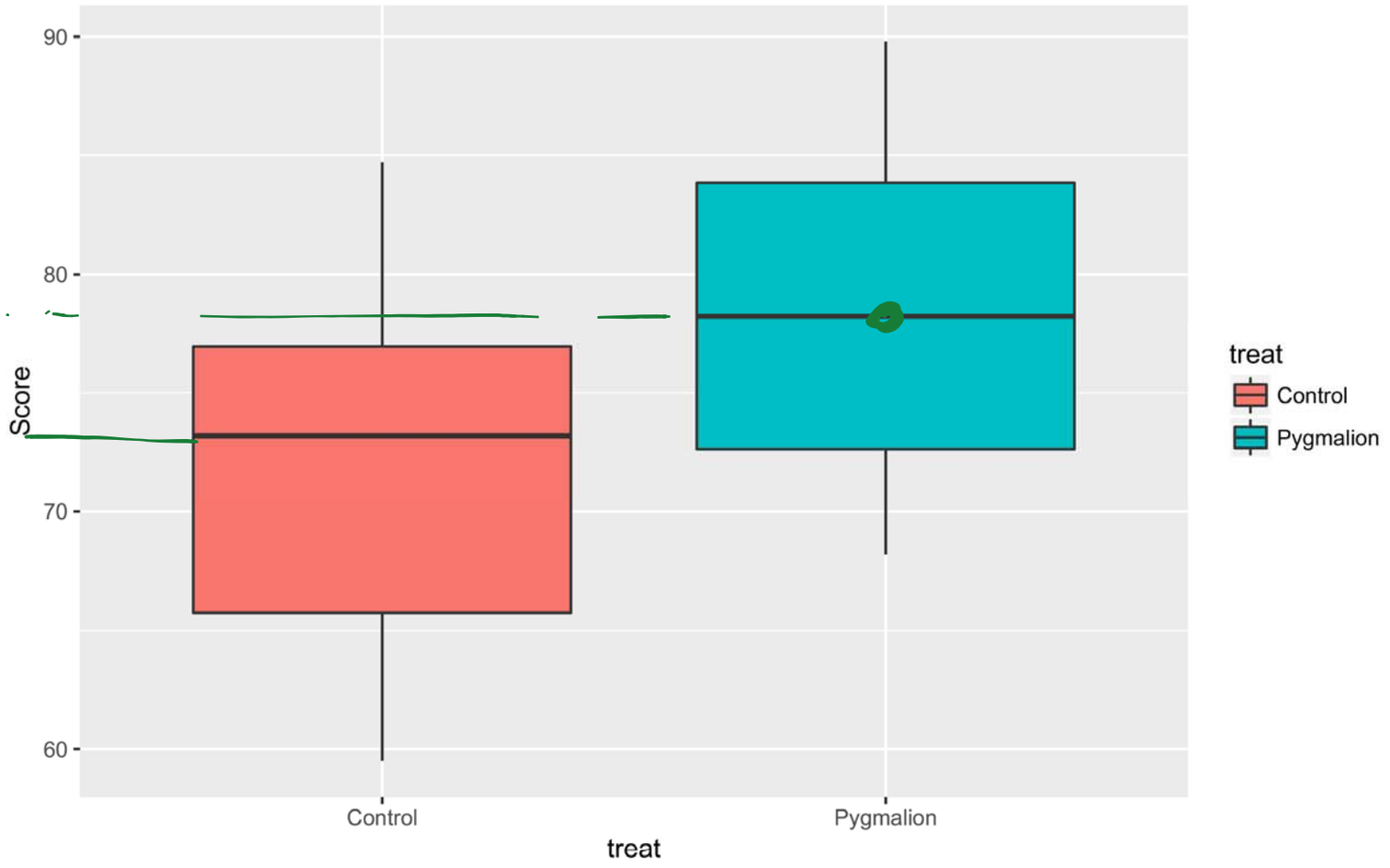
Case Study 2: Visualizing the data

```
#install.packages("ggplot2")
library(ggplot2)
pc<-ggplot(pyg, aes(x=company,y=Score, fill=company))+geom_boxplot()
pc
```



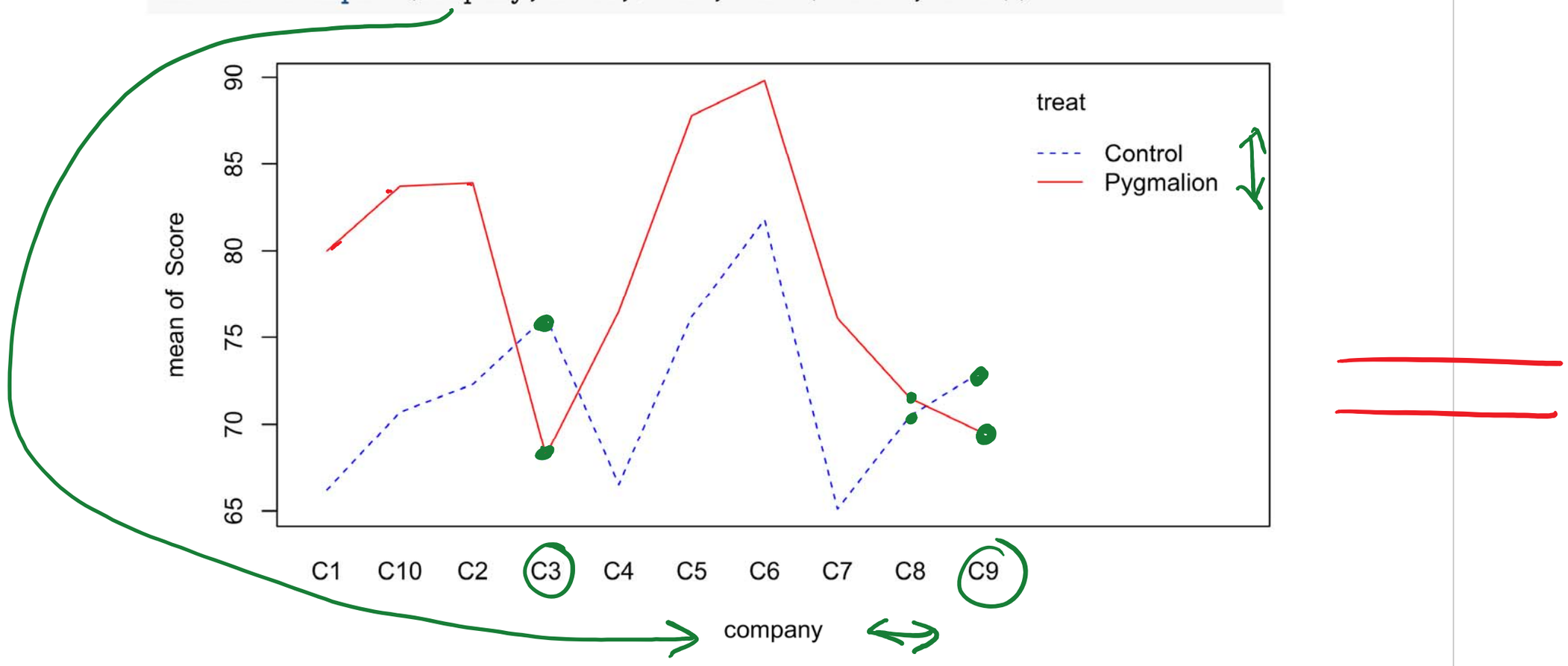
Case Study 2: Visualizing the data

```
ptr<-ggplot(pyg, aes(x=treat,y=Score, fill=treat))+geom_boxplot()
ptr
```



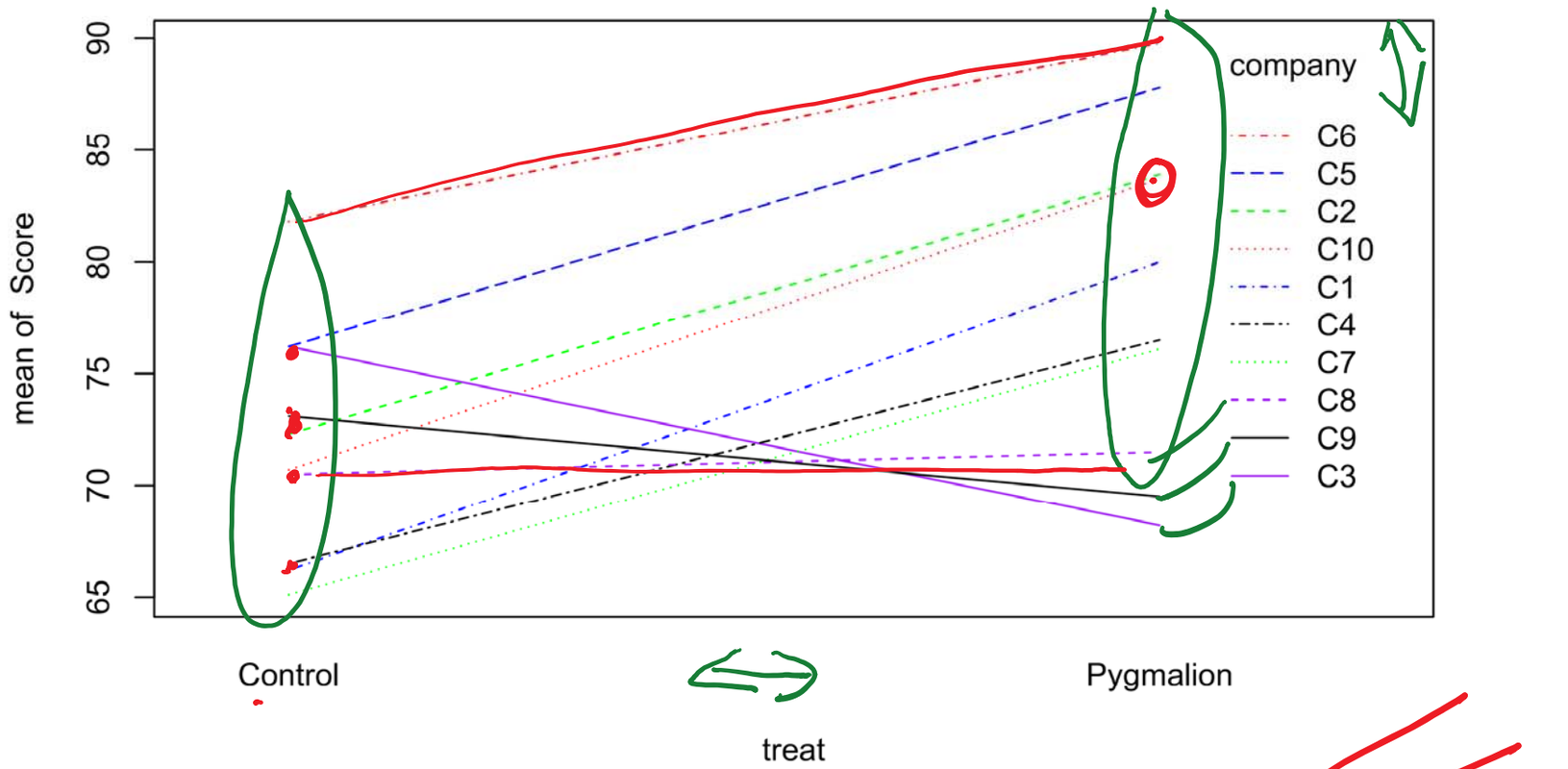
Case Study 2: Interaction plots

```
interaction.plot(company,treat,Score,col=c("blue","red"))
```



Case Study 2: Interaction plots

```
interaction.plot(treat,company,Score,col=c("blue", "red", "green","purple","bla
```



Case Study 2: Combination Means

```
cms=aggregate(Score~company+treat, data=pyg, FUN="mean")
cms[1:10,]
```

```
##    company   treat Score
## 1       C1 Control  66.2
## 2      C10 Control  70.7
## 3       C2 Control  72.3
## 4       C3 Control  76.2
## 5       C4 Control  66.5
## 6       C5 Control  76.2
## 7       C6 Control  81.8
## 8       C7 Control  65.1
## 9       C8 Control  70.5
## 10      C9 Control  73.1
```

Case Study 2: Combination Means

```
cms[11:20,]
```

```
##    company     treat Score
## 11      C1 Pygmalion  80.0
## 12     C10 Pygmalion  83.7
## 13      C2 Pygmalion  83.9
## 14      C3 Pygmalion  68.2
## 15      C4 Pygmalion  76.5
## 16      C5 Pygmalion  87.8
## 17      C6 Pygmalion  89.8
## 18      C7 Pygmalion  76.1
## 19      C8 Pygmalion  71.5
## 20      C9 Pygmalion  69.5
```

Case Study 2: Combination Means

```
tapply(Score, list(company,treat), mean)
```

```
##      Control Pygmalion
## C1      66.2      80.0
## C10     70.7      83.7
## C2      72.3      83.9
## C3      76.2      68.2
## C4      66.5      76.5
## C5      76.2      87.8
## C6      81.8      89.8
## C7      65.1      76.1
## C8      70.5      71.5
## C9      73.1      69.5
```

Case Study 2: Marginal Means

```
tapply(Score, company, mean)
```

```
##       C1      C10       C2       C3       C4       C5       C6       C7
## 70.80000 75.03333 76.16667 72.20000 69.83333 80.06667 84.46667 68.76667
##       C8       C9
## 70.83333 71.90000
```

```
tapply(Score, treat, mean)
```

```
##   Control Pygmalion
##  71.63158  78.70000
```

Case Study 2: Interaction model summary

```
##
## Call:
## lm(formula = Score ~ company * treat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -9.2   -2.3    0.0    2.3    9.2
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  66.200      5.094  12.996 3.89e-07 ***
## companyC10                    4.500      7.204   0.625   0.5477
## companyC2                     6.100      7.204   0.847   0.4191
## companyC3                    10.000      8.823   1.133   0.2863
## companyC4                     0.300      7.204   0.042   0.9677
## companyC5                    10.000      7.204   1.388   0.1985
## companyC6                    15.600      7.204   2.166   0.0585 .
## companyC7                    -1.100      7.204  -0.153   0.8820
## companyC8                     4.300      7.204   0.597   0.5653
## companyC9                     6.900      7.204   0.958   0.3632
## treatPygmalion               13.800      8.823   1.564   0.1522
## companyC10:treatPygmalion    -0.800     12.477  -0.064   0.9503
## companyC2:treatPygmalion     -2.200     12.477  -0.176   0.8639
## companyC3:treatPygmalion    -21.800     13.477  -1.618   0.1402
## companyC4:treatPygmalion     -3.800     12.477  -0.305   0.7676
## companyC5:treatPygmalion     -2.200     12.477  -0.176   0.8639
```

Case Study 2: Interaction model

```
anova(lm(Score~company*treat))
```

```
## Analysis of Variance Table
##
## Response: Score
##                Df Sum Sq Mean Sq F value  Pr(>F)
## company         9 670.98   74.55  1.4367 0.29902
## treat           1 338.88  338.88  6.5304 0.03092 *
## company:treat   9 311.46   34.61  0.6669 0.72212
## Residuals       9 467.04   51.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$SSReg_{full}$

$DFR_{full} = 9+1+9 = 19$

$DFE_{reduced} = 9$

DFE_{full}

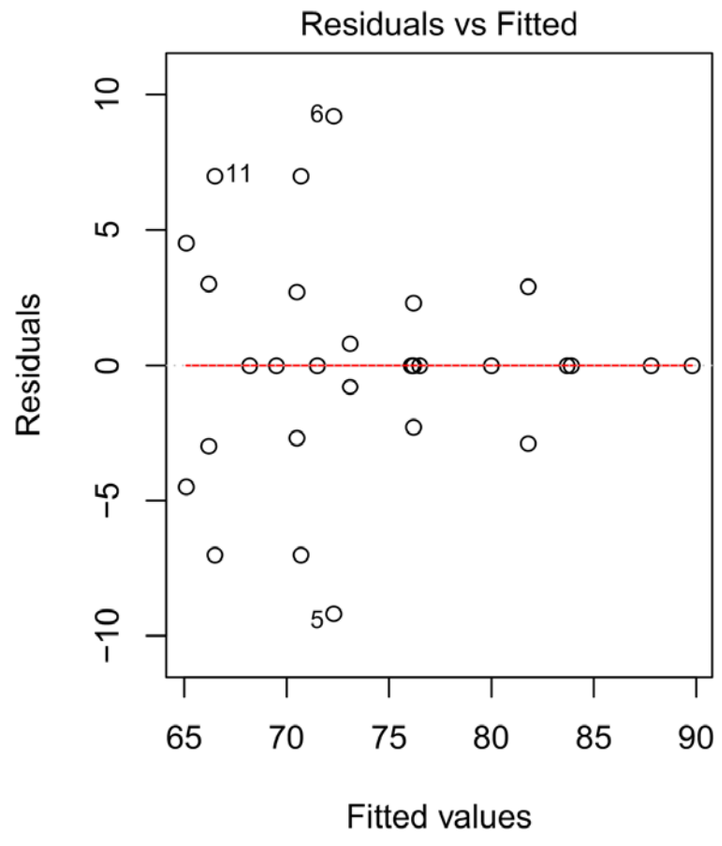
MSE_{full}

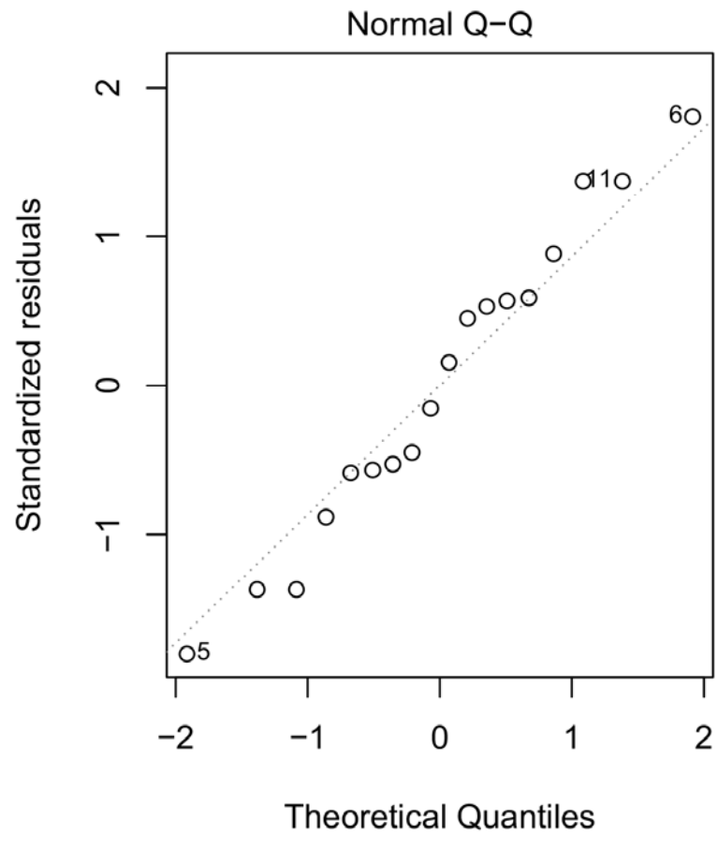
RSS_{full}

Case Study 2: Checking assumptions

```
fiti=lm(Score~company*treat, data=pyg)
par(mfrow=c(1,2))
plot(fiti, which=1:2)
```

```
## Warning: not plotting observations with leverage one:
##   1, 4, 7, 8, 9, 12, 15, 18, 21, 24, 27
```





Case Study 2: Additive model summary

```
summary(lm(Score~company+treat))
```

```
##
## Call:
## lm(formula = Score ~ company + treat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.660  -4.147   1.853   3.853   7.740
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68.39316    3.89308  17.568 8.92e-13 ***
## companyC10      4.23333    5.36968   0.788   0.4407
## companyC2       5.36667    5.36968   0.999   0.3308
## companyC3       0.19658    6.01886   0.033   0.9743
## companyC4      -0.96667    5.36968  -0.180   0.8591
## companyC5       9.26667    5.36968   1.726   0.1015
## companyC6      13.66667    5.36968   2.545   0.0203 *
## companyC7      -2.03333    5.36968  -0.379   0.7094
## companyC8       0.03333    5.36968   0.006   0.9951
## companyC9       1.10000    5.36968   0.205   0.8400
## treatPygmalion  7.22051    2.57951   2.799   0.0119 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Case Study 2: Additive model

(reduced) $y/x \sim$ company

last factor (full) $y/x \sim$ company + treat

```
anova(lm(Score~company+treat))
```

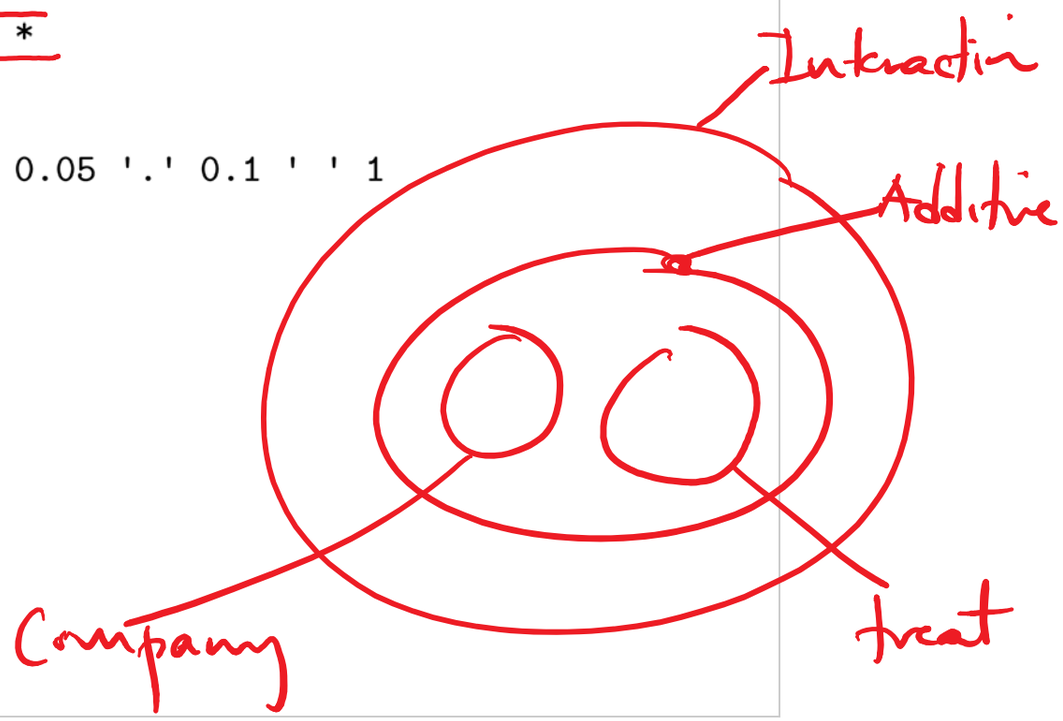
$DFR_{reduced} = 10$

$DFE_{reduced} = 18$

```
## Analysis of Variance Table
##
## Response: Score
##           Df Sum Sq Mean Sq F value  Pr(>F)
## company    9 670.98   74.55  1.7238 0.15556
## treat      1 338.88  338.88  7.8354 0.01186 *
## Residuals 18 778.50   43.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$SSReg_{reduced}$

$RSS_{reduced}$



Case Study 2: Additive model summary

```
summary(lm(Score~treat+company))
```

```
##
## Call:
## lm(formula = Score ~ treat + company)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.660  -4.147   1.853   3.853   7.740
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68.39316    3.89308  17.568 8.92e-13 ***
## treatPygmalion  7.22051    2.57951   2.799   0.0119 *
## companyC10      4.23333    5.36968   0.788   0.4407
## companyC2       5.36667    5.36968   0.999   0.3308
## companyC3       0.19658    6.01886   0.033   0.9743
## companyC4      -0.96667    5.36968  -0.180   0.8591
## companyC5       9.26667    5.36968   1.726   0.1015
## companyC6      13.66667    5.36968   2.545   0.0203 *
## companyC7      -2.03333    5.36968  -0.379   0.7094
## companyC8       0.03333    5.36968   0.006   0.9951
## companyC9       1.10000    5.36968   0.205   0.8400
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Case Study 2: Additive model

y/x ~ treat

y/x ~ treat + company

last

```
anova(lm(Score~treat+company))
```

$H_0: \beta_2 = \beta_3 = \cdots = \beta_{10} = 0$

```
## Analysis of Variance Table
##
## Response: Score
##           Df Sum Sq Mean Sq F value  Pr(>F)
## treat      1 327.34  327.34  7.5685 0.01314 *
## company    9 682.52   75.84  1.7534 0.14844
## Residuals 18 778.50   43.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

DFE_{full}

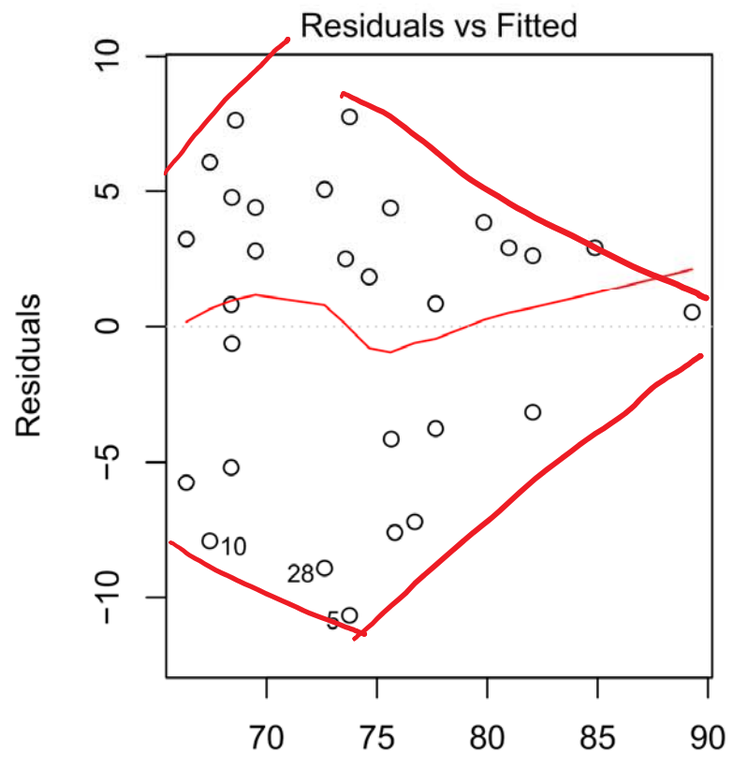
MSE_{full}

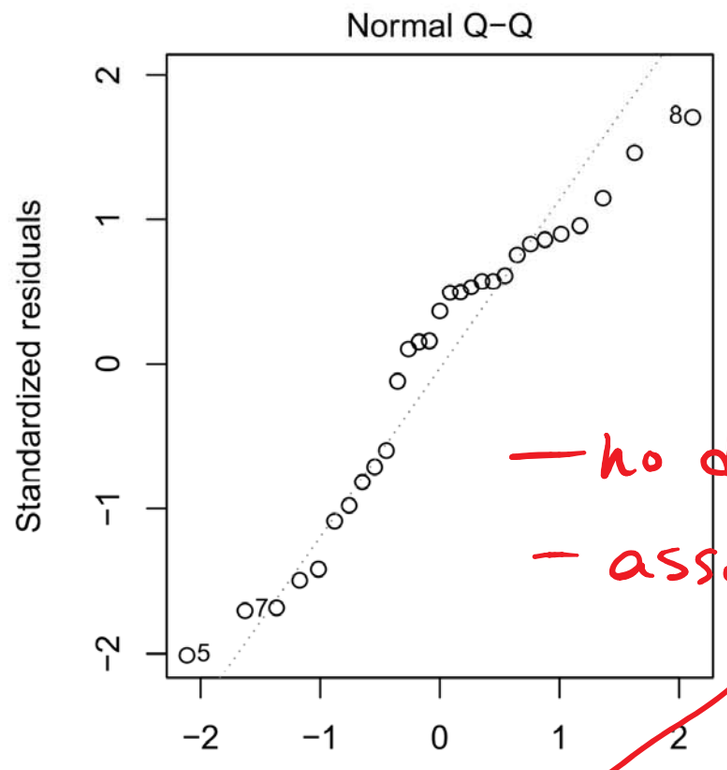
$F = 1.75 = \frac{75.84}{43.25}$

$= \frac{682.52/9}{43.25}$

Case Study 2: Checking assumptions

```
fita=lm(Score~company+treat, data=pyg)
par(mfrow=c(1,2))
plot(fita, which=1:2)
```





— no dramatic pattern
— assume normality.

— possibly the variance is decreasing
— consider weighted L.S. regression

Case Study 2: Saturated model as an ANOVA

```
fit1=aov(Score~company*treat, data=pyg)
summary(fit1)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## company         9  671.0    74.6   1.437 0.2990
## treat           1  338.9   338.9   6.530 0.0309 *
## company:treat   9  311.5    34.6   0.667 0.7221
## Residuals       9  467.0    51.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Case Study 2: Additive model as an ANOVA

```r
fit2=aov(Score~company+treat, data=pyg)
summary(fit2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## company      9  671.0    74.6   1.724 0.1556
## treat        1  338.9   338.9   7.835 0.0119 *
## Residuals   18  778.5    43.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Case Study 2:

```
t.crit=qt(1-0.05/2,df=27)
t.crit
```

```
## [1] 2.051831
```

```
spooled=sqrt((9*var(Score[treat=="Pygmalion"])+18*var(Score[treat=="Control"])),
spooled
```

```
## [1] 7.356078
```

```
t.test(Score[treat=="Pygmalion"], Score[treat=="Control"], var.equal=T)
```

```
##
##  Two Sample t-test
##
## data:  Score[treat == "Pygmalion"] and Score[treat == "Control"]
## t = 2.4595, df = 27, p-value = 0.0206
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.171707 12.965135
## sample estimates:
## mean of x mean of y
##  78.70000  71.63158
```