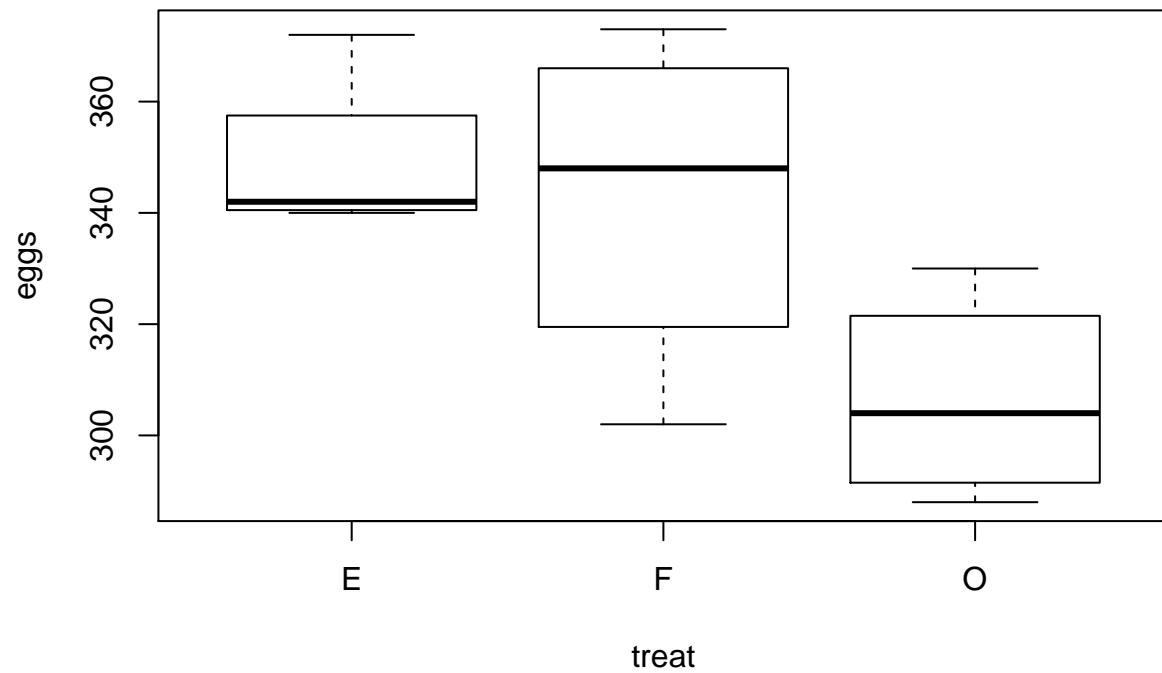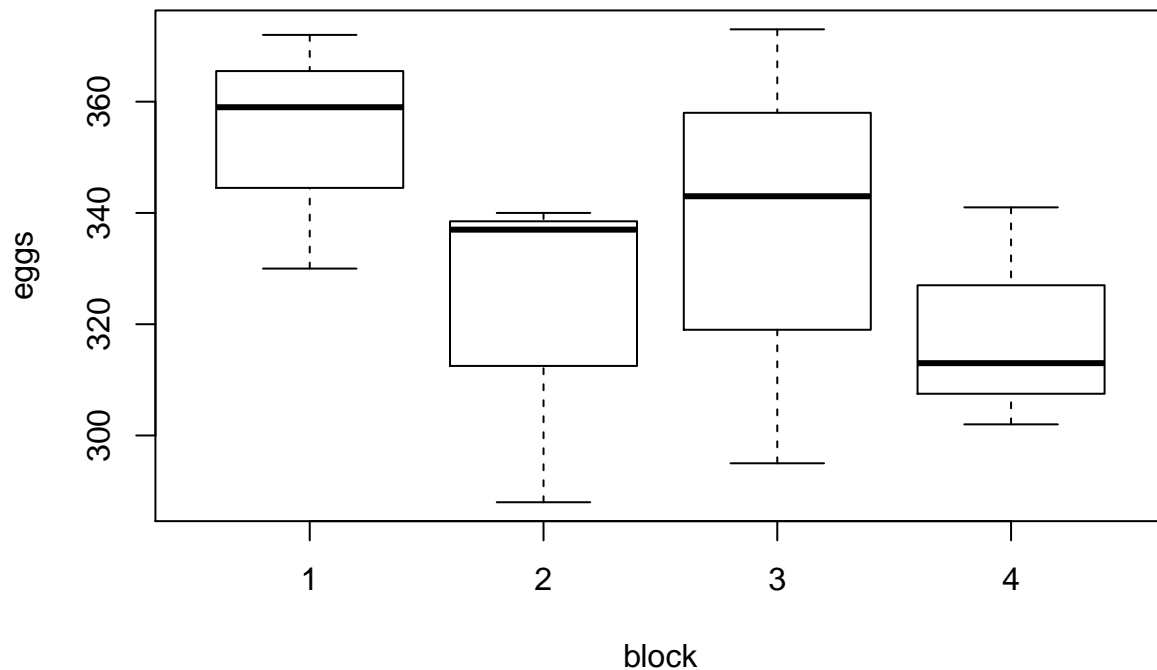# HW9

Nathan Horswill

April 11, 2019

Ch 10, q3 a

```
data(eggprod,package="faraway")
plot(eggs~treat,data=eggprod)
```



```
plot(eggs~block,data=eggprod)
```

While it is unclear if there is a difference in means for treatments E anf F, clearly treatment O differs significantly. Based on this I would anticipate the lowest egg production from block 2, treatment O, and the highest from block 1, treatment E or F.

3 b

```
lmod<-lm(eggs~treat+block,data=eggprod)
summary(lmod)
```

```
##
## Call:
## lm(formula = eggs ~ treat + block, data = eggprod)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.667  -8.125   2.083   5.521  26.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   369.92      13.91  26.596 1.87e-07 ***
## treatF         -6.25      13.91  -0.449   0.6690
## treatO        -42.50      13.91  -3.056   0.0224 *
## block2        -32.00      16.06  -1.992   0.0934 .
## block3        -16.67      16.06  -1.038   0.3394
## block4        -35.00      16.06  -2.179   0.0721 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.67 on 6 degrees of freedom
## Multiple R-squared:  0.7381, Adjusted R-squared:  0.5199
## F-statistic: 3.382 on 5 and 6 DF,  p-value: 0.08504
```

```r
par(mfrow=c(2,2))
plot(lmod)
```
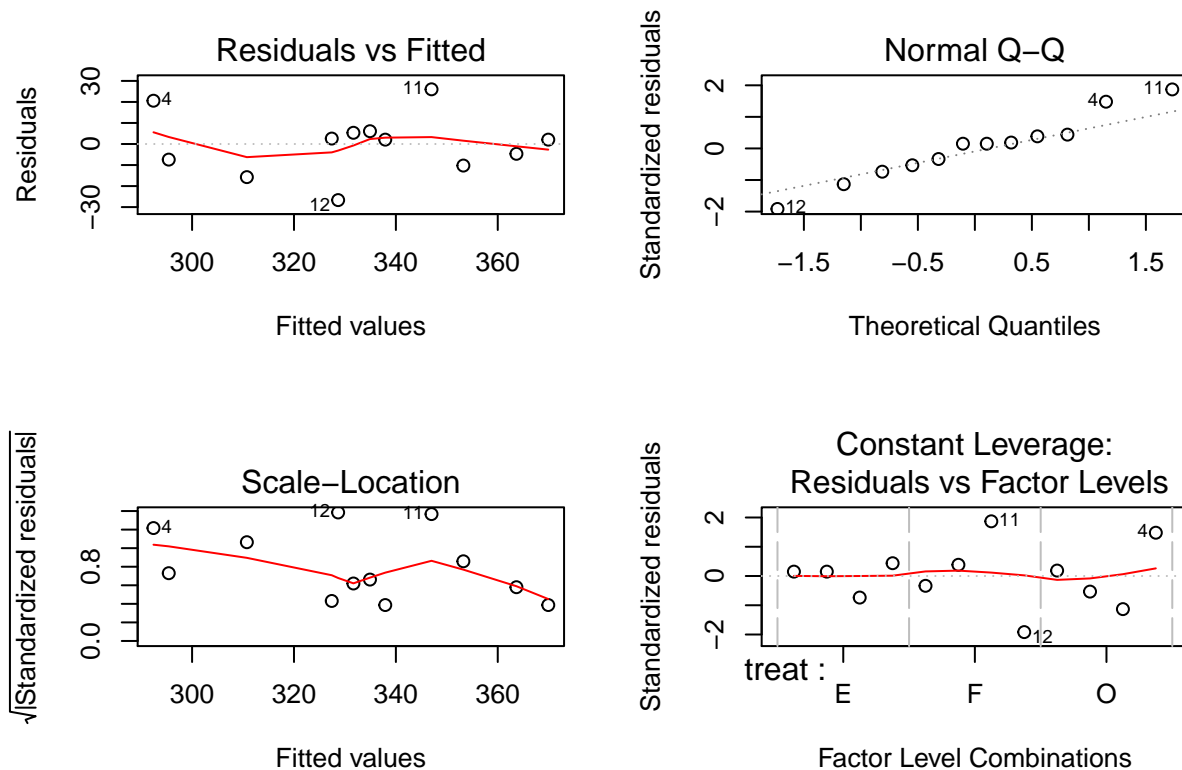

As we saw in the plots in part a, there is a more significant effect due to treatment than to block. There are no indications of an ill fitting model from the diagnostics plots.

3 c

```r
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.5.2
```

```
## Loading required package: Matrix
```

```r
library(faraway)
lrmod<-lmer(eggs~treat+(1|block),data=eggprod,REML=F)
summary(lrmod)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: eggs ~ treat + (1 | block)
##    Data: eggprod
##
```

```
##      AIC      BIC   logLik deviance df.resid
##    114.9    117.3    -52.4    104.9        7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.97723 -0.54794 -0.03286  0.74127  1.65055
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  block    (Intercept)  97.46    9.872
##  Residual             290.19   17.035
## Number of obs: 12, groups:  block, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  349.000      9.844  35.452
## treatF        -6.250     12.045  -0.519
## treatO       -42.500     12.045  -3.528
##
## Correlation of Fixed Effects:
##        (Intr) treatF
## treatF -0.612
## treatO -0.612  0.500
```

```
sumary(lrmod)
```

```
## Fixed Effects:
##             coef.est coef.se
## (Intercept) 349.00     9.84
## treatF       -6.25    12.05
## treatO      -42.50    12.05
##
## Random Effects:
##  Groups   Name        Std.Dev.
##  block    (Intercept)  9.87
##  Residual             17.03
## ---
## number of obs: 12, groups: block, 4
## AIC = 114.9, DIC = 104.9
## deviance = 104.9
```

According to the model treatment E results in the highest production of eggs. The standard error of treatment F is larger than the effect (as measured by it's coefficient), so treatment F is not statistically significantly different from treatment E.

3 d

```
library(pbkrtest)
ntmod<-lmer(eggs~1+(1|block),data=eggprod,REML=F)
```

```
## singular fit
```

```
KRmodcomp(lrmod,ntmod)
```

```
## F-test with Kenward-Roger approximation; computing time: 0.09 sec.
## large : eggs ~ treat + (1 | block)
```

```
## small : eggs ~ 1 + (1 | block)
##         stat    ndf    ddf F.scaling p.value
## Ftest 5.4437 2.0000 6.0000         1 0.04485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is clear indication of a significant difference in the treatments.

3 e

```
lrstat<-numeric(1000)
for (i in 1:1000){
  reggs<-unlist(simulate(ntmod))
  amod<-refit(lrmod,reggs)
  nmod<-refit(ntmod,reggs)
  lrstat[i]<-2*(logLik(amod,REML=F)-logLik(nmod,REML=F))
}
mean(lrstat>2*(logLik(lrmod,REML=F)-logLik(ntmod,REML=F)))
```

```
## [1] 0.066
```

Though the p-value acquired through bootstrapping is slightly higher than in part f, it is not significantly different. It does increase to over .05, so it appears to not achieve statistical significance. However, it is sufficiently close to the result form part (d) that is highlights how arbitrary the cuttoff of .05 is.

3 f

```
lrstat<-numeric(1000)
nkmod<-lm(eggs~1,data=eggprod)
ltmod<-lmer(eggs~1+(1|block),data=eggprod,REML=F)
stat<-2*(logLik(ltmod)-logLik(nkmod))
pv<-pchisq(stat,1,lower.tail = F)
for (i in 1:1000){
  seggs<-unlist(simulate(ntmod))
  amod<-lmer(seggs~1+(1|block),data=eggprod,REML=F)
  nmod<-lm(seggs~1,data=eggprod)
  lrstat[i]<-2*(logLik(amod,REML=F)-logLik(nmod,REML=F))
}
mean(lrstat>stat)
```
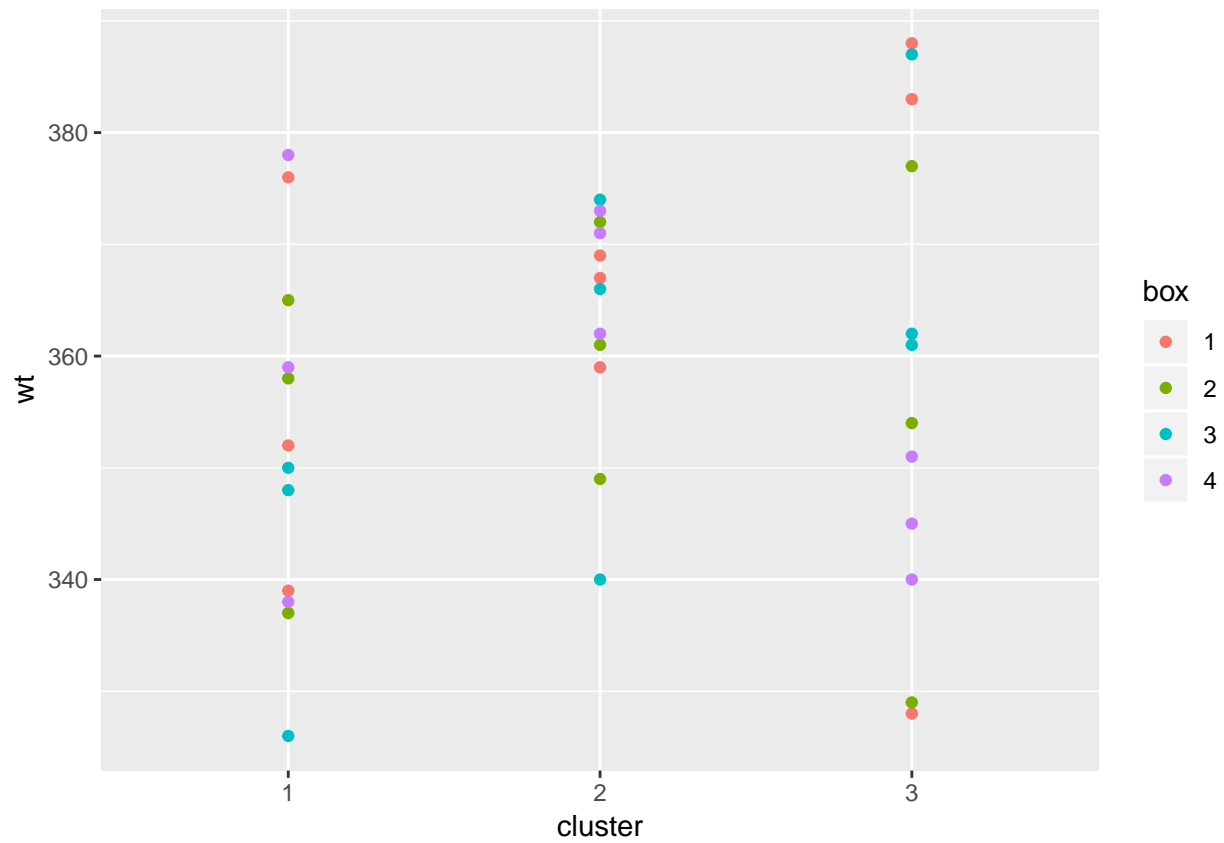
```
## [1] 0.663
```

We reject the hypothesis that there is a statistically significant effect due to block. The concurs with our model in part (b).

Ch 10, q5 a

```
library(ggplot2)
```
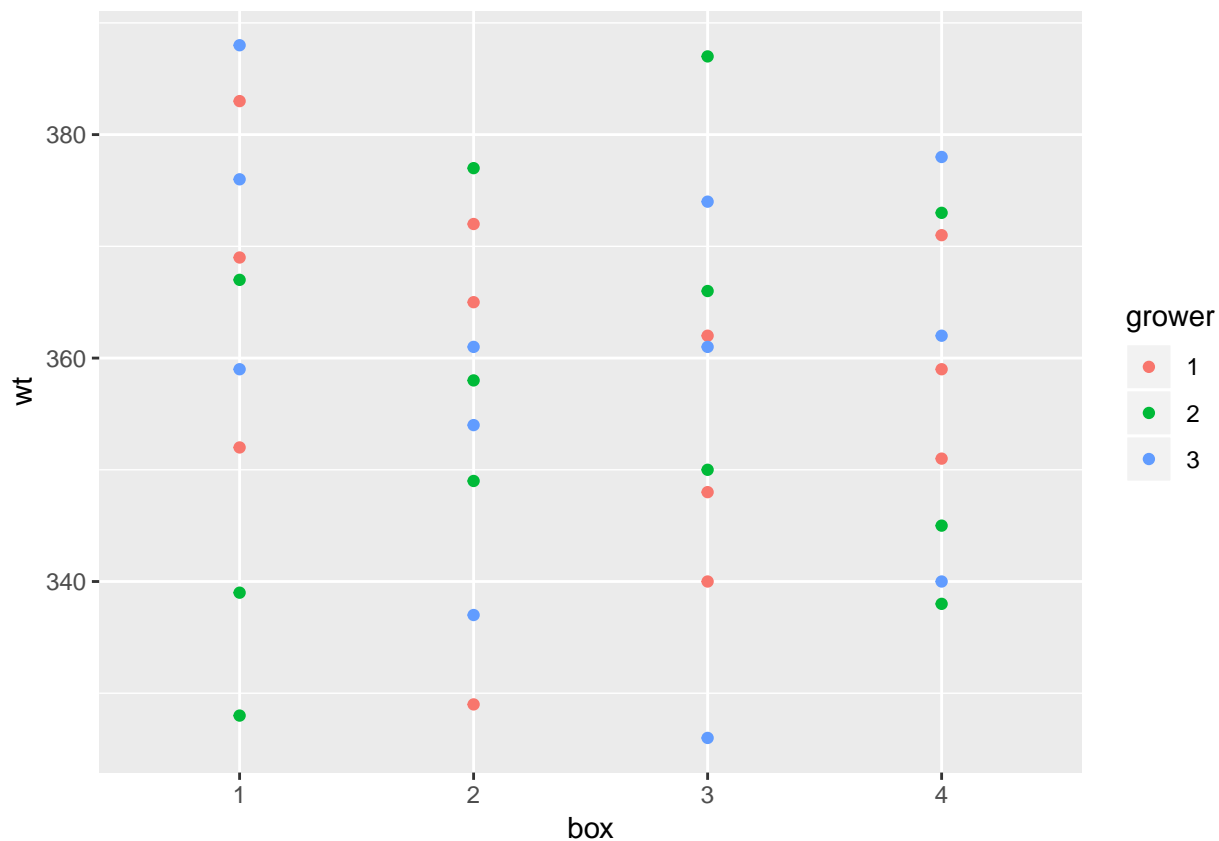
```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
data("broccoli")
ggplot(data=broccoli,aes(x=cluster,y=wt,col=box))+geom_point()
```

```
ggplot(data=broccoli,aes(x=box,y=wt,col=grower))+geom_point()
```

Through the above 3 graphs it is very difficult to see if there is a statistically significant difference between clusters, bozes or growers. We anticipate very minimal effect when we control for these three variables in a model.

5 b

```
library(tidyverse)
broccoli%>%group_by(grower)%>%summarise(average = mean(wt),stddev=sd(wt))

## # A tibble: 3 x 3
##   grower average stddev
##   <fct>    <dbl>  <dbl>
## 1 1         358.   15.2
## 2 2         356.   17.9
## 3 3         360.   18.3

broccoli%>%group_by(box)%>%summarise(average = mean(wt),stddev=sd(wt))

## # A tibble: 4 x 3
##   box   average stddev
##   <fct>   <dbl>  <dbl>
## 1 1        362.   19.9
## 2 2        356.   15.6
## 3 3        357.   18.3
## 4 4        357.   14.8
```

We can see the standard deviations of each grouping is far larger than the group differences. This confirms

our intuition that there is not a statistically significant differences between group means.

5 c

```
remod<-lmer(wt~cluster+grower+(1|box),data=broccoli,REML=F)
```

```
## singular fit
```

```
summary(remod)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: wt ~ cluster + grower + (1 | box)
##    Data: broccoli
##
##      AIC      BIC   logLik deviance df.resid
##    314.7    325.8   -150.3    300.7       29
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.90369 -0.73371  0.01058  0.53012  1.90369
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  box      (Intercept)   0.0      0.00
##  Residual             248.3     15.76
## Number of obs: 36, groups:  box, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  352.417      5.873  60.006
## cluster2      11.417      6.434   1.775
## cluster3       6.583      6.434   1.023
## grower2       -2.000      6.434  -0.311
## grower3        1.250      6.434   0.194
##
## Correlation of Fixed Effects:
##         (Intr) clstr2 clstr3 growr2
## cluster2 -0.548
## cluster3 -0.548  0.500
## grower2  -0.548  0.000  0.000
## grower3  -0.548  0.000  0.000  0.500
## convergence code: 0
## singular fit
```

None of the t values above show significance of the fixed effects.

5 d

```
grmod<-lmer(wt~(1|box)+cluster,data=broccoli,REML=F)
```

```
## singular fit
```

```
KRmodcomp(grmod,remod)
```

```
## singular fit
```

```
## F-test with Kenward-Roger approximation; computing time: 0.08 sec.
## large : wt ~ cluster + grower + (1 | box)
## small : wt ~ (1 | box) + cluster
##          stat     ndf     ddf F.scaling p.value
## Ftest  0.1118  2.0000 28.0000         1  0.8946
```

There is very little evidence of a statistically significant effect due to grower.

5 e

```
amod<-lmer(wt~(1|box)+cluster+grower,data=broccoli,REML=F)
bmod<-lm(wt~grower+cluster,data=broccoli)
lrstat<-numeric(1000)
stat<-2*(logLik(amod)-logLik(bmod))
pv<-pchisq(stat,1,lower.tail = F)
c(stat,pv)
```

```
## [1] 0 1
```

```
for (i in 1:1000){
  swt<-unlist(simulate(bmod))
  amod<-lmer(swt~(1|box)+cluster+grower,data=broccoli,REML=F)
  nmod<-lm(swt~grower+cluster,data=broccoli)
  lrstat[i]<-2*(logLik(amod,REML=F)-logLik(nmod,REML=F))
}

mean(lrstat>stat)
```

```
## [1] 0.561
```

We fail to see evidence that the boxed effect is significant.

5 f

```
confint(remod,method="boot")
```

```
##                    2.5 %      97.5 %
## .sig01        0.00000000    6.426823
## .sigma       10.48948440   18.169124
## (Intercept) 341.09899383  364.122758
## cluster2     -0.09593006   24.342408
## cluster3     -5.57231576   18.633696
## grower2     -14.33477459    9.838431
## grower3     -12.01280344   13.966313
```