HW 7

Ted Henson

3/30/2020

```r
knitr::opts_chunk$set(message=F, warning=F,  eval = T)
```
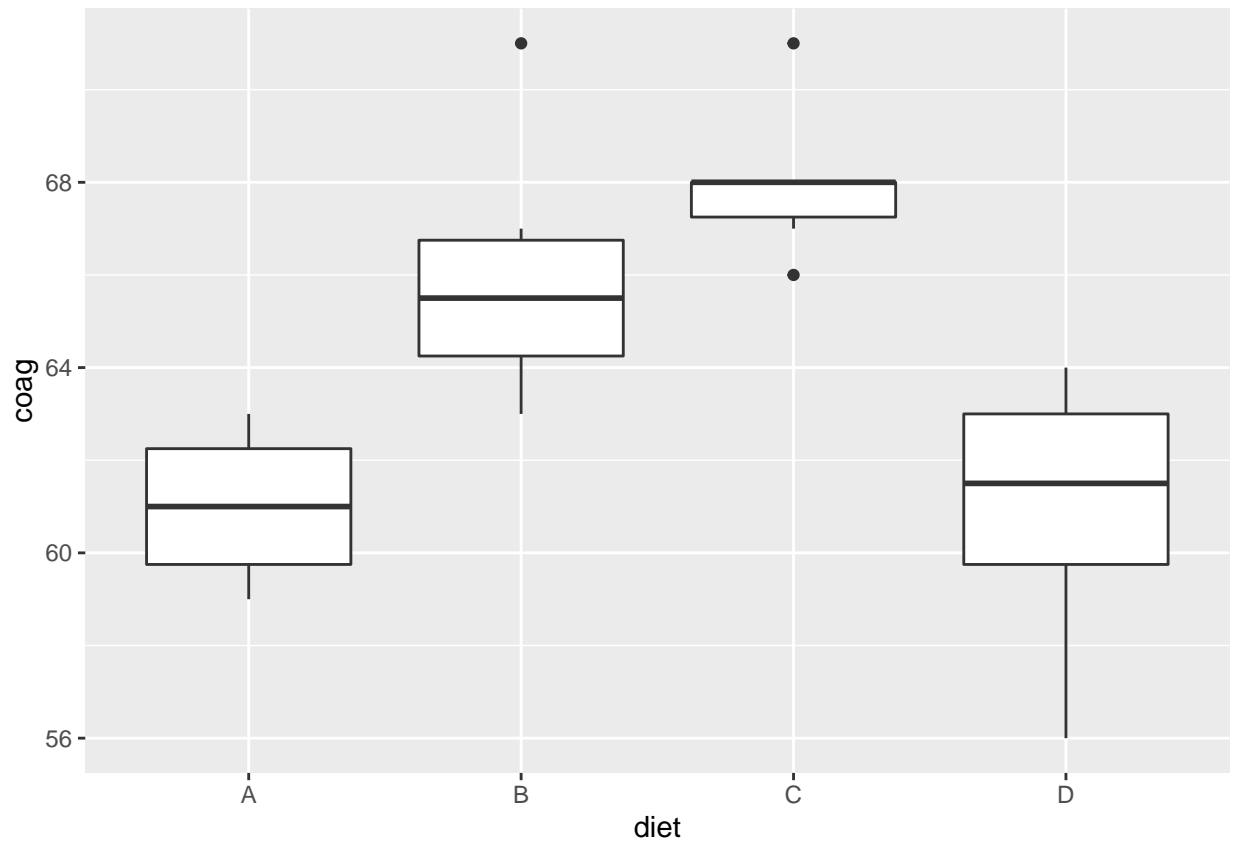
Question 2

a)

```r
library(faraway)
data(coagulation)
head(coagulation)
```

```
##   coag diet
## 1   62    A
## 2   60    A
## 3   63    A
## 4   59    A
## 5   63    B
## 6   67    B
```

```r
library(ggplot2)
library(tidyverse)
ggplot(coagulation, aes(x = diet, y = coag)) + geom_boxplot()
```

It appears that diets D and A have low coagulation times comapred to diets B and C, which have very high coagulation times, especially C.

b)

```
op=options(contrasts=c('contr.sum','contr.poly'))
lmod=aov(coag~diet,coagulation)
summary(lmod)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## diet         3    228    76.0   13.57 4.66e-05 ***
## Residuals   20    112     5.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
new.data = data.frame(diet='D')
predict(lmod, new.data, interval = 'prediction')
```

```
##   fit      lwr      upr
## 1  61 55.76427 66.23573
```

c)

```
library(lme4)
mmod=lmer(coag~diet+(1|diet),coagulation)
options(op)
```

```r
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: coag ~ diet + (1 | diet)
##    Data: coagulation
##
## REML criterion at convergence: 101
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1129 -0.5282  0.0000  0.5282  2.1129
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  diet     (Intercept) 34.88    5.906
##  Residual              5.60    2.366
## Number of obs: 24, groups:  diet, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   64.000      2.995  21.371
## diet1         -3.000      5.207  -0.576
## diet2          2.000      5.184   0.386
## diet3          4.000      5.184   0.772
##
## Correlation of Fixed Effects:
##       (Intr) diet1  diet2
## diet1  0.007
## diet2 -0.001 -0.335
## diet3 -0.001 -0.335 -0.333
```

```r
new.data = data.frame(diet='D')
fit = predict(mmod, new.data,se.fit=T)

group.sd=as.data.frame(VarCorr(mmod))$sdcor[1]
resid.sd=as.data.frame(VarCorr(mmod))$sdcor[2]
pv=numeric(1000)
for(i in 1:1000){
y=unlist(simulate(mmod))
bmod=suppressMessages(refit(mmod,y))
pv[i]=predict(bmod,new.data=data.frame(operator='D'))+rnorm(n=1,sd=resid.sd)
}
quantile(pv,c(0.025,0.975))
```

```
##     2.5%    97.5%
## 49.03126 73.61417
```

d)

```r
for(i in 1:1000){
y=unlist(simulate(mmod))
bmod=suppressMessages(refit(mmod,y))
pv[i]=predict(bmod,re.form=~0)[1]+rnorm(n=1,sd=group.sd)+rnorm(n=1,sd=resid.sd)
}
```

```
quantile(pv,c(0.025,0.975))
```

```
##     2.5%    97.5%
## 44.34056 77.88616
```
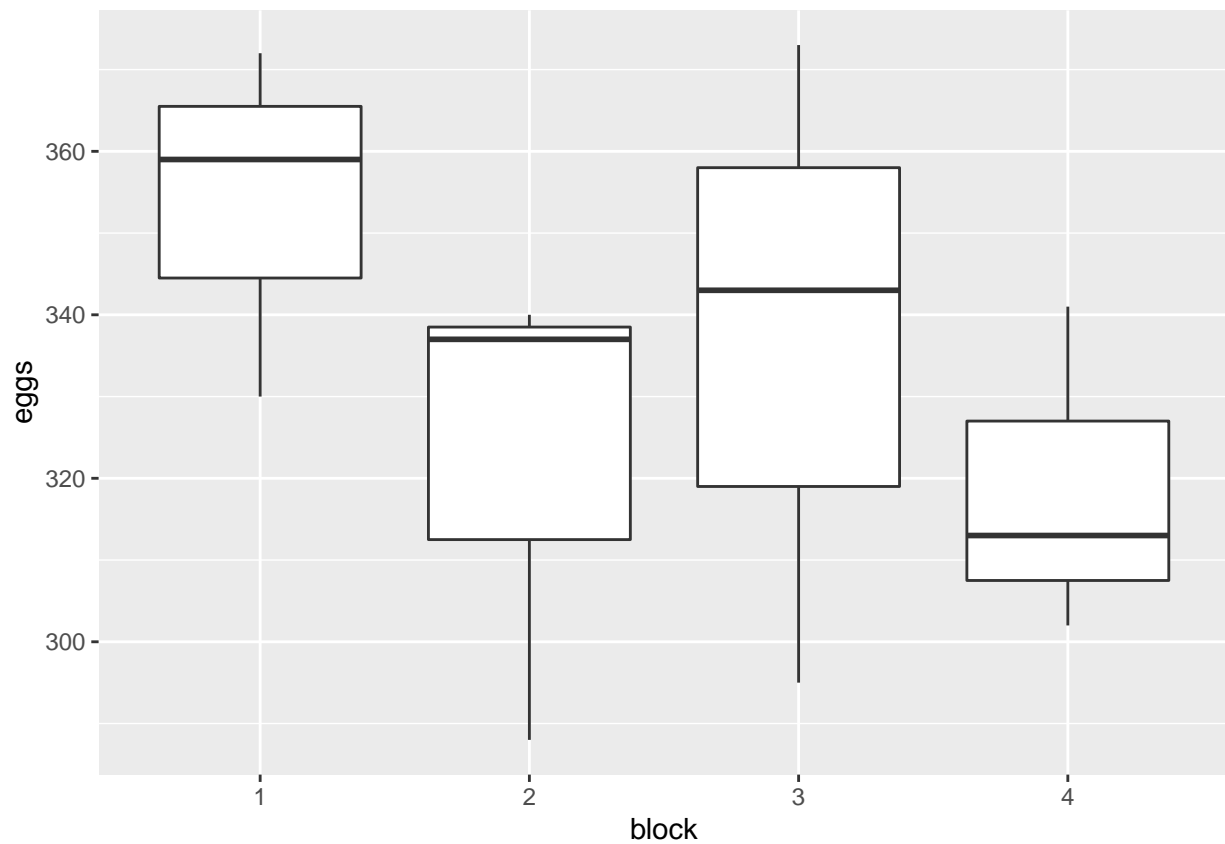
e)

```
for(i in 1:1000){
y=unlist(simulate(mmod))
bmod=suppressMessages(refit(mmod,y))
pv[i]=predict(bmod,re.form=~0)[1]+rnorm(n=1,sd=group.sd)+rnorm(n=1,sd=resid.sd)
}
quantile(pv,c(0.025,0.975))
```

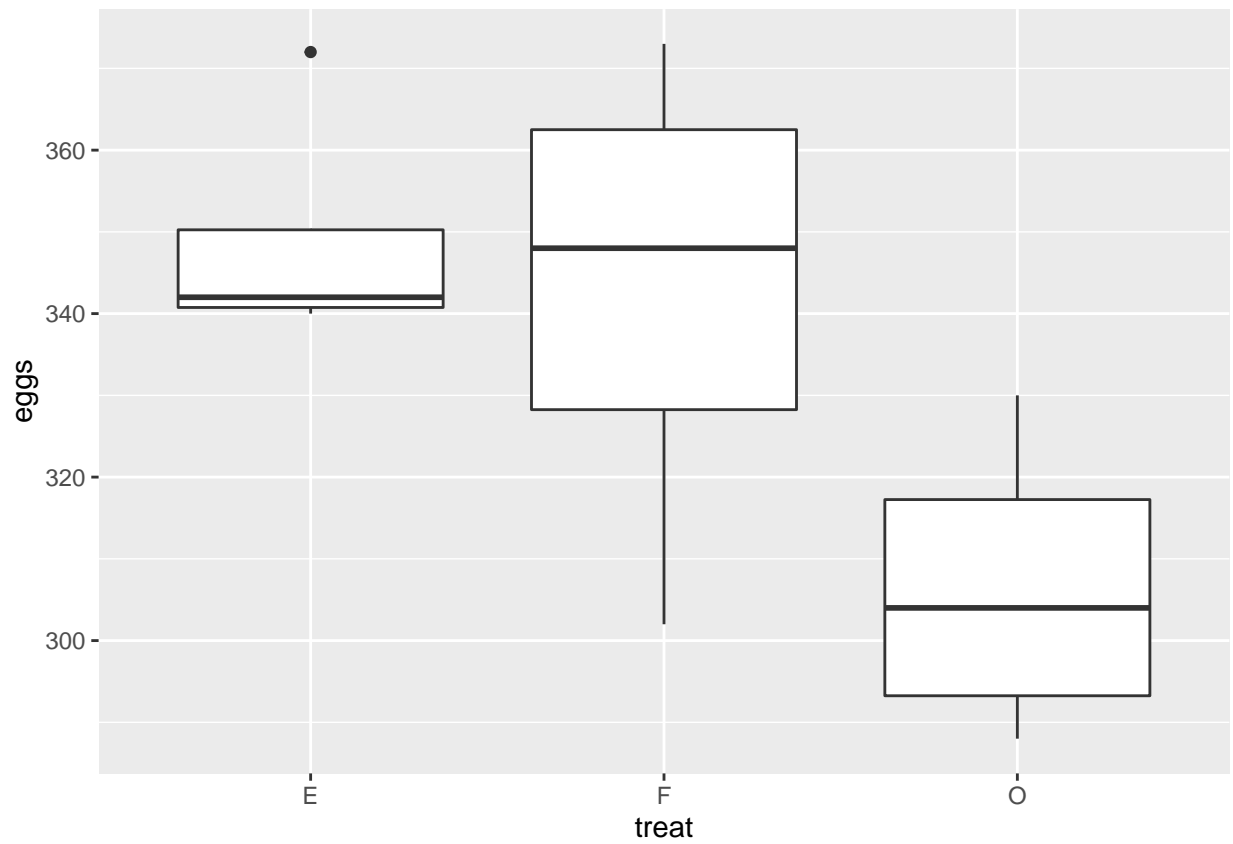```
2.5%   97.5%
44.54016 77.34417
```

# Question 3
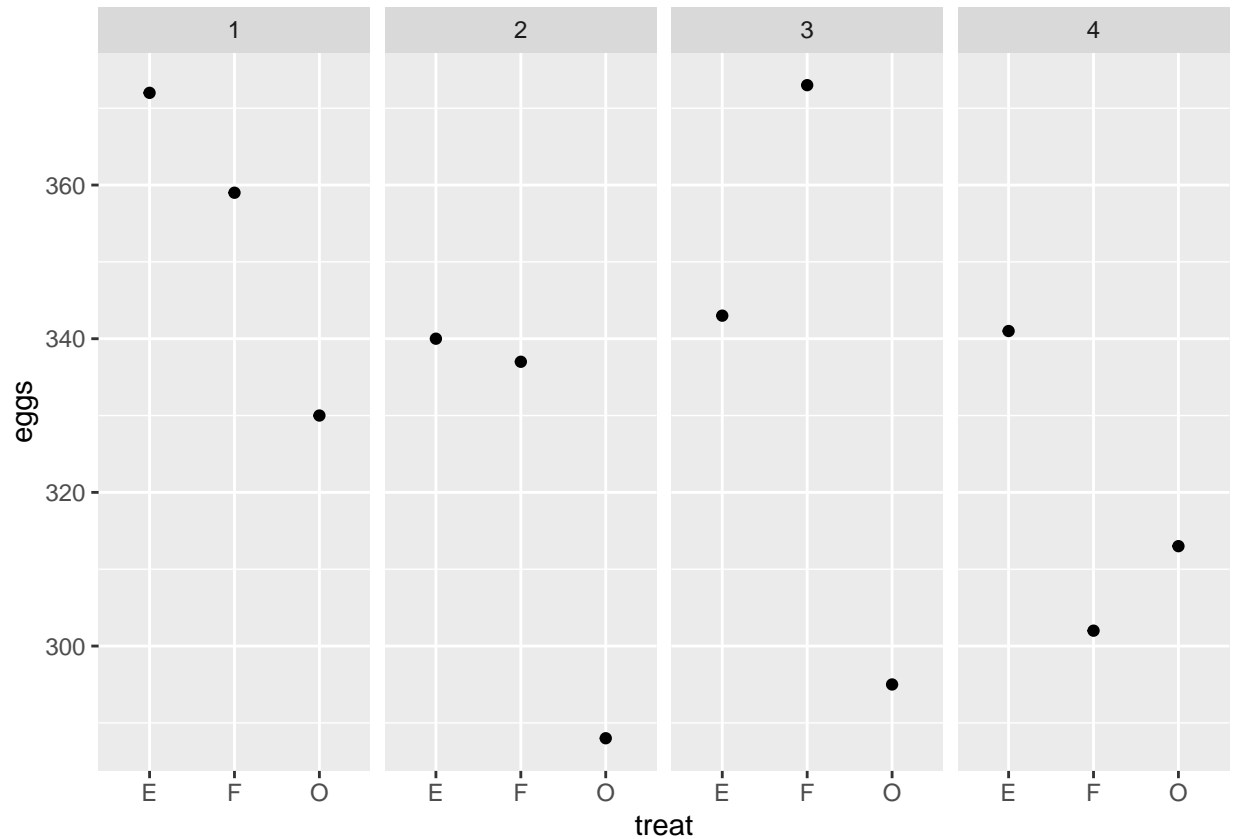
a)

```
data(eggprod)
ggplot(eggprod, aes(x=block, y = eggs)) + geom_boxplot()
```

```r
ggplot(eggprod, aes(x=treat, y = eggs)) + geom_boxplot()
```



```r
ggplot(eggprod, aes(x=treat, y = eggs)) + geom_point() + facet_grid(~ block)
```

Across the board treatment o and block 4 appear to produce the lowest eggs, but when combined, these effects do not remain. The lowest number of eggs occured with treatment O and block 2.

b)

```
lmod=aov(eggs~treat+block,eggprod)
summary(lmod)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## treat        2   4212  2106.2   5.444 0.0449 *
## block        3   2330   776.8   2.008 0.2145
## Residuals    6   2322   386.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the p values the treatment variable apperas to be significant, but the block variable may not be after accounting for the treatment.

c)

```
mmod=lmer(eggs~treat+(1|block),eggprod)
summary(mmod)
```

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

6

```
## 
## REML criterion at convergence: 85.4
## 
## Scaled residuals:
##      Min       1Q    Median       3Q      Max 
## -1.71233 -0.47454 -0.02843  0.64192  1.42943 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  block    (Intercept) 130.0    11.40   
##  Residual             386.9    19.67   
## Number of obs: 12, groups:  block, 4
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   349.00      11.37  30.702
## treatF         -6.25      13.91  -0.449
## treatO        -42.50      13.91  -3.056
## 
## Correlation of Fixed Effects:
##        (Intr) treatF
## treatF -0.612       
## treatO -0.612  0.500
```

```
fixef(mmod)
```

```
## (Intercept)       treatF       treatO 
##      349.00        -6.25       -42.50
```

According to this model, treatment E is best at maximizing egg production, but we cannot be sure because of the random effects imposed by the block variable.

d)

```
library(pbkrtest)
nmod=lmer(eggs~1+(1|block),eggprod)
```

```
KRmodcomp(mmod,nmod)
```

```
## F-test with Kenward-Roger approximation; computing time: 0.20 sec.
## large : eggs ~ treat + (1 | block)
## small : eggs ~ 1 + (1 | block)
##         stat    ndf    ddf F.scaling p.value
## Ftest 5.4438 2.0000 6.0000         1 0.04485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Kenward-Roger approximation agrees with the fixed effects model that the treatments are significant.

e)

```
#parametric bootstrap
amod=lmer(eggs~treat+(1|block),eggprod,REML=F)
nmod=lmer(eggs~1+(1|block),eggprod,REML=F)
pmod=PBmodcomp(amod,nmod)
```

```
summary(pmod)
```

```
## Parametric bootstrap test; time: 30.24 sec; samples: 1000 extremes: 64;
## large : eggs ~ treat + (1 | block)
## small : eggs ~ 1 + (1 | block)
##            stat     df    ddf p.value
## PBtest   8.4245                0.06494 .
## Gamma    8.4245                0.05869 .
## Bartlett 5.6075 2.0000         0.06058 .
## F        4.2123 2.0000 2.9976 0.13466
## LRT      8.4245 2.0000         0.01481 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The paramatric bootstrap agrees that the treatments are significant.

f)

```
rmod=lmer(eggs~treat+(1|block),eggprod)
nlmod=lm(eggs~treat,eggprod)
as.numeric(2*(logLik(rmod)-logLik(nlmod,REML=T)))
```

```
## [1] 0.5153561
```

```
lrtstatf=numeric(1000)
for(i in 1:1000){
reggs=unlist(simulate(nlmod))
nlmodr=lm(reggs~treat,eggprod)
#rmodr=refit(rmod,reggs)
rmodr=lmer(reggs~treat+(1|block),eggprod)
lrtstatf[i]=2*(logLik(rmodr)-logLik(nlmodr,REML=T))
}

mean(lrtstatf<0.00001)
```

```
## [1] 0.575
```

```
mean(lrtstatf>as.numeric(2*(logLik(rmod)-logLik(nlmod,REML=T))))
```

```
## [1] 0.212
```

The bootstrapping test agrees with the fixed results test that the blocks are not significant.