Assignment 7 KEY

36 points total, 2 points per problem part unless otherwise noted.

PCB (2x2 factorial)

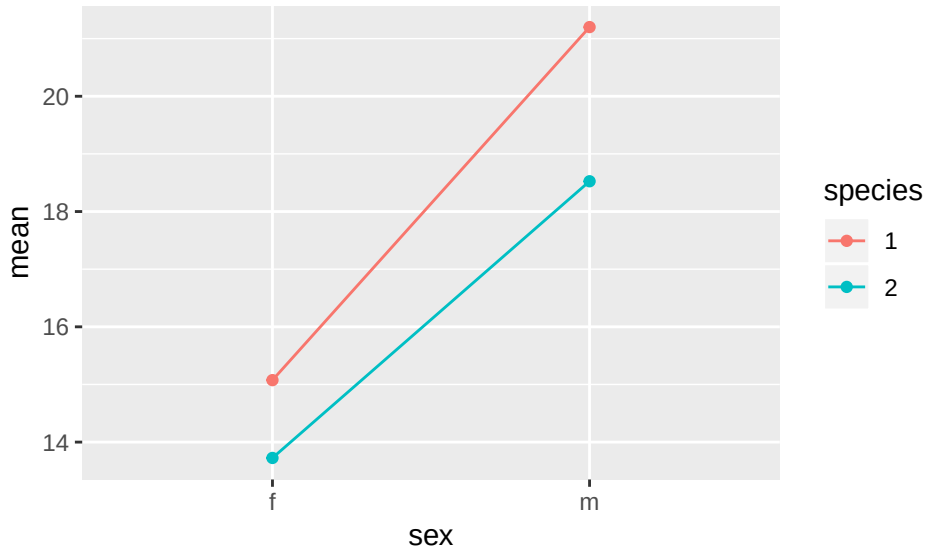
```
library(dplyr)
library(ggplot2)
library(car)
library(emmeans)
PCBdata <- read.csv("~/Dropbox/STAT512/Assigns/Assign7/PCB.csv")
#str(PCBdata)
PCBdata$species <- as.factor(PCBdata$species)
```

A. Summary Statistics and Graph (**4pts**)

```
SumStats <- summarize(group_by(PCBdata, sex, species),
                      n = n(),
                      mean = mean(pcb),
                      sd = sd(pcb))
SumStats
```

```
## # A tibble: 4 x 5
## # Groups:   sex [2]
##   sex   species     n  mean    sd
##   <fct> <fct>   <int> <dbl> <dbl>
## 1 f     1           4  15.1  1.24
## 2 f     2           4  13.7  1.21
## 3 m     1           4  21.2  1.33
## 4 m     2           4  18.5  1.84
```

```
qplot(x = sex, y = mean, group = species, color = species, data = SumStats) +
  geom_line()
```



B. One-way ANOVA table

```
options(contrasts=c("contr.sum","contr.poly"))
Model1 <- lm(pcb ~ group, data = PCBdata)
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: pcb
##             Sum Sq Df  F value    Pr(>F)
## (Intercept) 4695.7  1 2309.112 4.317e-15 ***
## group        137.3  3   22.508 3.230e-05 ***
## Residuals     24.4 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C. One-way pairwise comparisons

```
emmeans(Model1, pairwise ~ group)
```

```
## $emmeans
##  group emmean        SE df lower.CL upper.CL
##  sp1f  15.075 0.7130115 12 13.52148 16.62852
##  sp1m  21.200 0.7130115 12 19.64648 22.75352
##  sp2f  13.725 0.7130115 12 12.17148 15.27852
##  sp2m  18.525 0.7130115 12 16.97148 20.07852
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast    estimate       SE df t.ratio p.value
##  sp1f - sp1m   -6.125 1.008351 12  -6.074  0.0003
##  sp1f - sp2f    1.350 1.008351 12   1.339  0.5576
##  sp1f - sp2m   -3.450 1.008351 12  -3.421  0.0227
##  sp1m - sp2f    7.475 1.008351 12   7.413  <.0001
##  sp1m - sp2m    2.675 1.008351 12   2.653  0.0857
##  sp2f - sp2m   -4.800 1.008351 12  -4.760  0.0023
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

D. Two-way ANOVA table

```
Model2 <- lm(pcb ~ sex*species, data = PCBdata)
Anova(Model2, type =3)
```

```
## Anova Table (Type III tests)
##
## Response: pcb
##             Sum Sq Df   F value    Pr(>F)
## (Intercept) 4695.7  1 2309.1121 4.317e-15 ***
## sex          119.4  1   58.6935 5.839e-06 ***
## species       16.2  1    7.9667   0.01539 *
## sex:species    1.8  1    0.8633   0.37112
## Residuals     24.4 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E. Two-way pairwise comparisons (Interaction)

```
emmeans(Model2, pairwise ~ sex*species)
```

```
## $emmeans
##  sex species emmean        SE df lower.CL upper.CL
##  f   1       15.075 0.7130115 12 13.52148 16.62852
##  m   1       21.200 0.7130115 12 19.64648 22.75352
##  f   2       13.725 0.7130115 12 12.17148 15.27852
##  m   2       18.525 0.7130115 12 16.97148 20.07852
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast  estimate       SE df t.ratio p.value
##  f,1 - m,1   -6.125 1.008351 12  -6.074  0.0003
##  f,1 - f,2    1.350 1.008351 12   1.339  0.5576
##  f,1 - m,2   -3.450 1.008351 12  -3.421  0.0227
##  m,1 - f,2    7.475 1.008351 12   7.413  <.0001
##  m,1 - m,2    2.675 1.008351 12   2.653  0.0857
##  f,2 - m,2   -4.800 1.008351 12  -4.760  0.0023
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

F. Two-way pairwise comparisons (Species Main Effect)

```
emmeans(Model2, pairwise ~ species)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##  species  emmean        SE df lower.CL upper.CL
##  1       18.1375 0.5041753 12  17.0390  19.2360
##  2       16.1250 0.5041753 12  15.0265  17.2235
##
## Results are averaged over the levels of: sex
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate        SE df t.ratio p.value
##  1 - 2      2.0125 0.7130115 12   2.823  0.0154
##
## Results are averaged over the levels of: sex
```

G. The SE is smaller for the main effect comparison corresponding to species. We have higher power corresponding to the main effect comparisons because we are averaging over the other factor.

2. Roadway damage 2X3 Factorial with Blocking

```
library(dplyr)
library(ggplot2)
library(car)
library(emmeans)
options(contrasts=c("contr.sum", "contr.poly"))
InData <- read.csv("~/Dropbox/STAT512/Assigns/Assign7/Ex15-14.csv", header = TRUE)
str(InData)
```

```
## 'data.frame':    30 obs. of  4 variables:
##  $ Roadway      : int  1 1 1 1 1 1 2 2 2 2 ...
##  $ cracks       : int  37 49 43 47 27 33 39 50 42 48 ...
##  $ Treatment    : Factor w/ 3 levels "CaCl","NaCl",..: 2 2 1 1 3 3 2 2 1 1 ...
##  $ Concentration: Factor w/ 2 levels "High","Low": 2 1 2 1 2 1 2 1 2 1 ...
```

```
InData$Roadway <- as.factor(InData$Roadway)
```

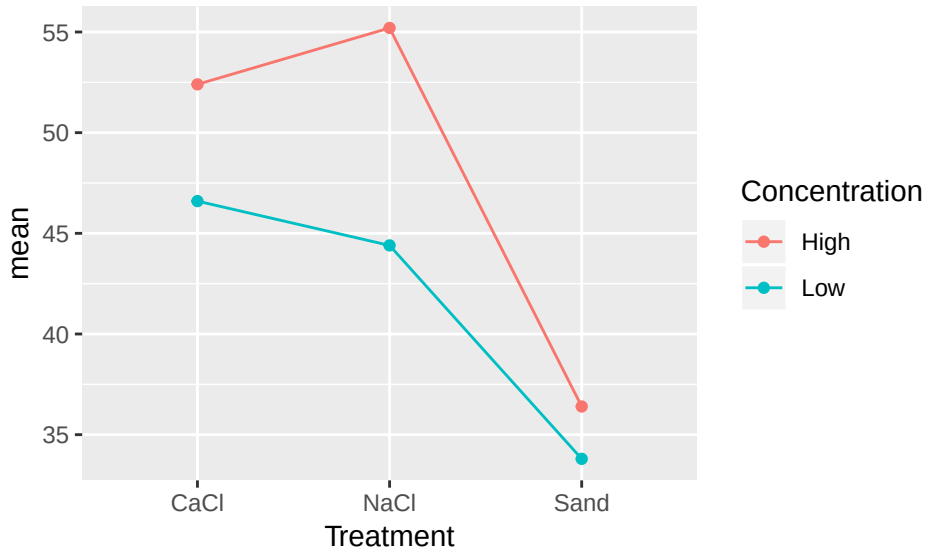
A. Blocking structure: Randomized Complete Bock.
Treatment Structure: 2-way factorial
Concentration (High,Low) by Treatment (CaCL, NaCl, Sand)

B.

```
SumStats <- summarize(group_by(InData, Treatment, Concentration),
                      n    = n(),
                      mean = mean(cracks),
                      sd   = sd(cracks))
SumStats
```

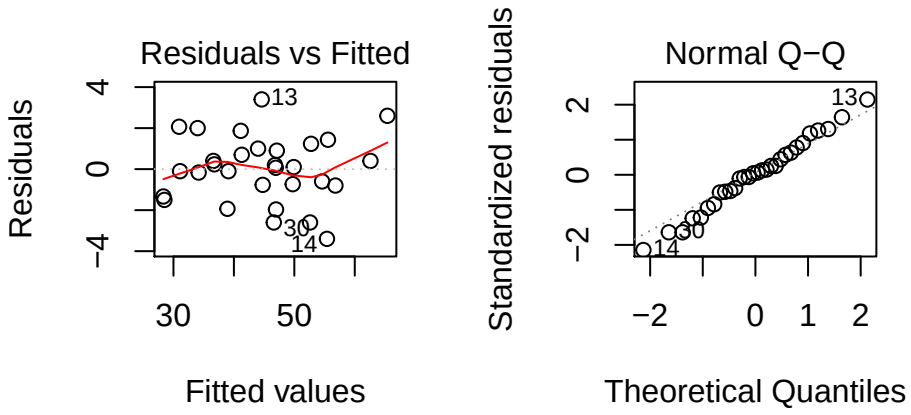
```
## # A tibble: 6 x 5
## # Groups:   Treatment [3]
##   Treatment Concentration     n  mean    sd
##   <fct>     <fct>         <int> <dbl> <dbl>
## 1 CaCl      High              5  52.4  6.50
## 2 CaCl      Low               5  46.6  5.59
## 3 NaCl      High              5  55.2  7.79
## 4 NaCl      Low               5  44.4  6.88
## 5 Sand      High              5  36.4  4.98
## 6 Sand      Low               5  33.8  7.46
```

```
qplot(x = Treatment, y = mean, colour = Concentration, group = Concentration, data = SumStats) + geom_li
```



C. The Diagnostics plot were not requested, but its a good idea to take a look at them. They'll be requested on the final.

```
RoadModel <- lm(cracks ~ Roadway + Treatment*Concentration, data = InData)
par(mfrow = c(1,2))
plot(RoadModel, which =1);plot(RoadModel, which =2)
```




```
Anova(RoadModel, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: cracks
##                          Sum Sq Df   F value    Pr(>F)
## (Intercept)               60211  1 15999.433 < 2.2e-16 ***
## Roadway                     973  4    64.646 3.740e-11 ***
## Treatment                  1412  2   187.573 1.103e-13 ***
## Concentration               307  1    81.630 1.694e-08 ***
## Treatment:Concentration      85  2    11.346 0.0005091 ***
## Residuals                    75 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D. Blocking was effective ($F = 64.462$, $p < 0.001$)

E.

```
emmeans(RoadModel, pairwise ~ Concentration|Treatment)
```

```
## $emmeans
## Treatment = CaCl:
##  Concentration emmean        SE df lower.CL upper.CL
##  High            52.4 0.8675636 20 50.59029 54.20971
##  Low             46.6 0.8675636 20 44.79029 48.40971
##
## Treatment = NaCl:
##  Concentration emmean        SE df lower.CL upper.CL
##  High            55.2 0.8675636 20 53.39029 57.00971
##  Low             44.4 0.8675636 20 42.59029 46.20971
##
## Treatment = Sand:
##  Concentration emmean        SE df lower.CL upper.CL
##  High            36.4 0.8675636 20 34.59029 38.20971
##  Low             33.8 0.8675636 20 31.99029 35.60971
##
## Results are averaged over the levels of: Roadway
## Confidence level used: 0.95
##
## $contrasts
## Treatment = CaCl:
```

```
##  contrast   estimate      SE df t.ratio p.value
##  High - Low      5.8 1.22692 20   4.727  0.0001
##
## Treatment = NaCl:
##  contrast   estimate      SE df t.ratio p.value
##  High - Low     10.8 1.22692 20   8.803  <.0001
##
## Treatment = Sand:
##  contrast   estimate      SE df t.ratio p.value
##  High - Low      2.6 1.22692 20   2.119  0.0468
##
## Results are averaged over the levels of: Roadway
```

Average number of cracks for High is significantly higher than Low for each of the 3 Treatments. F.

```
emmeans(RoadModel, pairwise ~ Treatment|Concentration)
```

```
## $emmeans
## Concentration = High:
##  Treatment emmean        SE df lower.CL upper.CL
##  CaCl        52.4 0.8675636 20 50.59029 54.20971
##  NaCl        55.2 0.8675636 20 53.39029 57.00971
##  Sand        36.4 0.8675636 20 34.59029 38.20971
##
## Concentration = Low:
##  Treatment emmean        SE df lower.CL upper.CL
##  CaCl        46.6 0.8675636 20 44.79029 48.40971
##  NaCl        44.4 0.8675636 20 42.59029 46.20971
##  Sand        33.8 0.8675636 20 31.99029 35.60971
##
## Results are averaged over the levels of: Roadway
## Confidence level used: 0.95
##
## $contrasts
## Concentration = High:
##  contrast    estimate      SE df t.ratio p.value
##  CaCl - NaCl     -2.8 1.22692 20  -2.282  0.0817
##  CaCl - Sand     16.0 1.22692 20  13.041  <.0001
##  NaCl - Sand     18.8 1.22692 20  15.323  <.0001
##
## Concentration = Low:
##  contrast    estimate      SE df t.ratio p.value
##  CaCl - NaCl      2.2 1.22692 20   1.793  0.1974
##  CaCl - Sand     12.8 1.22692 20  10.433  <.0001
##  NaCl - Sand     10.6 1.22692 20   8.640  <.0001
##
## Results are averaged over the levels of: Roadway
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Discussion: At Low concentration, average number of cracks for Sand is significantly lower than both CaCl and NaCl. There is not a statistically significant difference between CaCl and NaCl.