HW7 KEY

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

PCB (2x2 factorial)

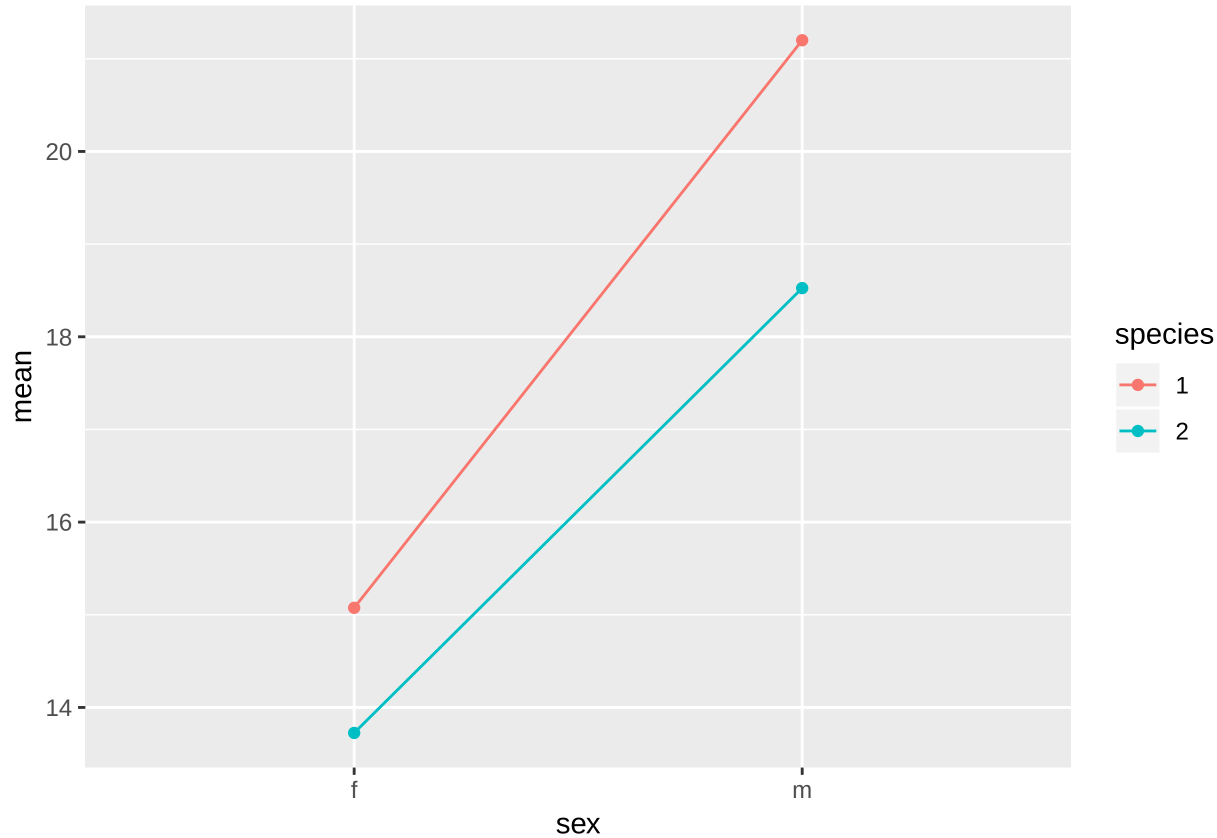
```r
library(dplyr)
library(ggplot2)
library(car)
library(emmeans)
PCBdata <- read.csv("C:/hess/STAT512/HW_2019/HW7/PCB.csv")
#str(PCBdata)
PCBdata$species <- as.factor(PCBdata$species)
```

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

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

```
## # A tibble: 4 x 5
## # Groups:   sex [?]
##   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
```

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



2. 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
```

3. One-way pairwise comparisons

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

```
## $emmeans
##  group emmean    SE df lower.CL upper.CL
##  sp1f    15.1 0.713 12     13.5     16.6
##  sp1m    21.2 0.713 12     19.6     22.8
##  sp2f    13.7 0.713 12     12.2     15.3
##  sp2m    18.5 0.713 12     17.0     20.1
##
## Confidence level used: 0.95
```

```
##
## $contrasts
##  contrast    estimate   SE df t.ratio p.value
##  sp1f - sp1m    -6.12 1.01 12 -6.074  0.0003
##  sp1f - sp2f     1.35 1.01 12  1.339  0.5576
##  sp1f - sp2m    -3.45 1.01 12 -3.421  0.0227
##  sp1m - sp2f     7.47 1.01 12  7.413  <.0001
##  sp1m - sp2m     2.67 1.01 12  2.653  0.0857
##  sp2f - sp2m    -4.80 1.01 12 -4.760  0.0023
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

4. Two-way ANOVA table

```r
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
```

5. Two-way pairwise comparisons (Interaction)

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

```
## $emmeans
##  sex species emmean    SE df lower.CL upper.CL
##  f   1         15.1 0.713 12     13.5     16.6
##  m   1         21.2 0.713 12     19.6     22.8
##  f   2         13.7 0.713 12     12.2     15.3
##  m   2         18.5 0.713 12     17.0     20.1
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast  estimate   SE df t.ratio p.value
##  f,1 - m,1    -6.12 1.01 12 -6.074  0.0003
##  f,1 - f,2     1.35 1.01 12  1.339  0.5576
##  f,1 - m,2    -3.45 1.01 12 -3.421  0.0227
##  m,1 - f,2     7.47 1.01 12  7.413  <.0001
##  m,1 - m,2     2.67 1.01 12  2.653  0.0857
##  f,2 - m,2    -4.80 1.01 12 -4.760  0.0023
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

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

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

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

```
## $emmeans
##  species emmean    SE df lower.CL upper.CL
##  1         18.1 0.504 12       17     19.2
##  2         16.1 0.504 12       15     17.2
##
## Results are averaged over the levels of: sex
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate    SE df t.ratio p.value
##  1 - 2        2.01 0.713 12 2.823   0.0154
##
## Results are averaged over the levels of: sex
```

7. 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. In addition a Tukey adjustment is used (by default) with the interaction comparisons this will also contribute to the larger p-value for interaction comparison.

Roadways (2x3 factorial with blocking)

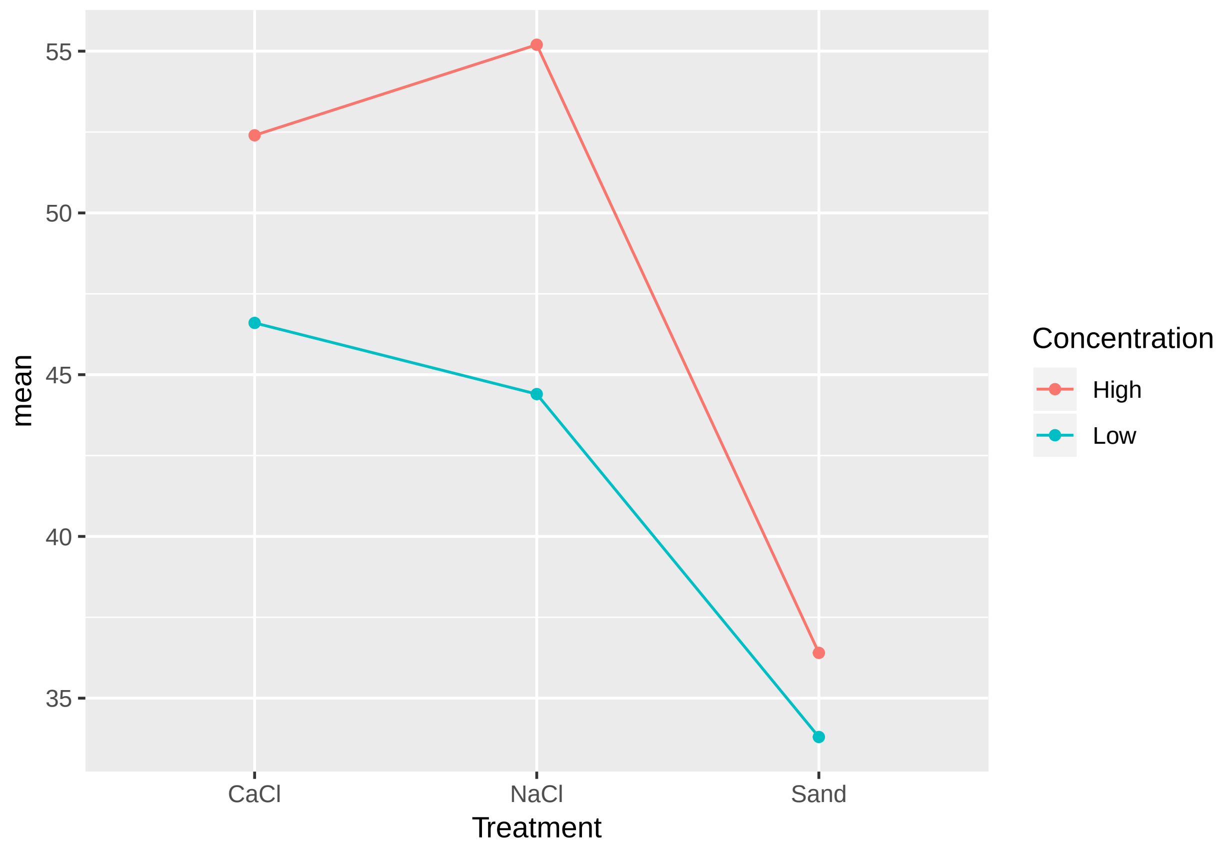
```
library(dplyr)
library(ggplot2)
library(car)
library(emmeans)
InData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH15/ex15-14.txt", quote = " ' ")
InData$Roadway <- as.factor(InData$Roadway)
```

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

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

```
## # A tibble: 6 x 5
## # Groups:   Treatment [?]
##   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_line()
```



9. The blocking structure is **RCB** (randomized complete block design). The treatment structure is **2x3 factorial**.
10. ANOVA Table (**4pts**)

```
options(contrasts=c("contr.sum", "contr.poly"))
RoadModel <- lm(cracks ~ Roadway + Treatment*Concentration, data = InData)
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
```

11. Blocking was effective ($F = 64.462$, $p < 0.001$).
12. Pairwise comparisons #1 (**4pts**)
    Average number of cracks for High is significantly higher than Low for each of the 3 Treatments.

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

```
## Treatment = CaCl:
##  contrast   estimate   SE df t.ratio p.value
##  High - Low      5.8 1.23 20 4.727   0.0001
##
## Treatment = NaCl:
##  contrast   estimate   SE df t.ratio p.value
##  High - Low     10.8 1.23 20 8.803   <.0001
##
## Treatment = Sand:
##  contrast   estimate   SE df t.ratio p.value
##  High - Low      2.6 1.23 20 2.119   0.0468
##
## Results are averaged over the levels of: Roadway
```

13. Pairwise comparisons #2 (**4pt**)
    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.

```r
emmeans(RoadModel, pairwise ~ Treatment|Concentration)$contrasts
```

```
## Concentration = High:
##  contrast    estimate   SE df t.ratio p.value
##  CaCl - NaCl     -2.8 1.23 20 -2.282  0.0817
##  CaCl - Sand     16.0 1.23 20 13.041  <.0001
##  NaCl - Sand     18.8 1.23 20 15.323  <.0001
##
## Concentration = Low:
##  contrast    estimate   SE df t.ratio p.value
##  CaCl - NaCl      2.2 1.23 20  1.793  0.1974
##  CaCl - Sand     12.8 1.23 20 10.433  <.0001
##  NaCl - Sand     10.6 1.23 20  8.640  <.0001
##
## Results are averaged over the levels of: Roadway
## P value adjustment: tukey method for comparing a family of 3 estimates
```