STAA 553: HW6

YOUR NAME HERE

See Canvas Calendar for due date.
40 points total, 2 points per problem unless otherwise noted.
Content for all questions is from section 10 or earlier.

Important Notes:
- Remember to check `str()` and then define things `as.factor` where needed.
- A key feature of the balanced designs considered in this assignment are that the SE's are the same for all means and pairwise comparisons.

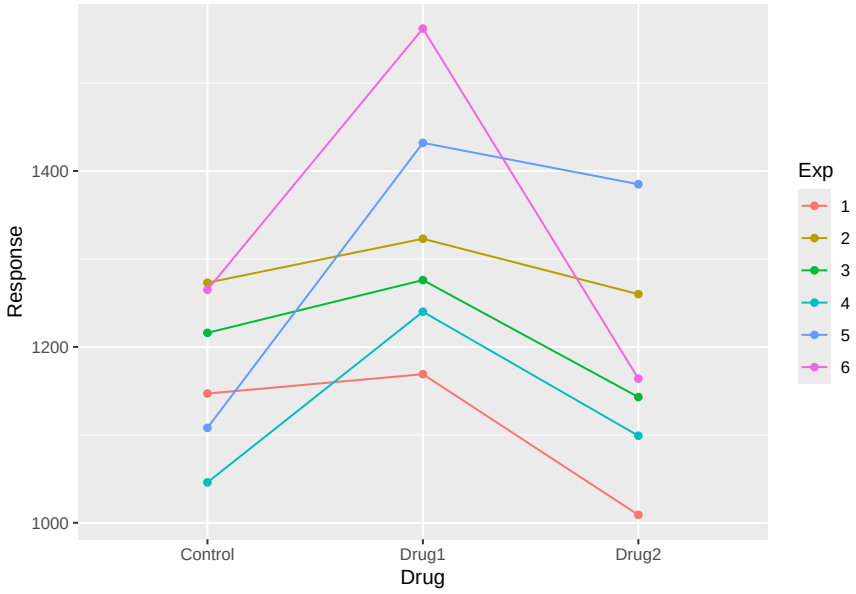
Cell Line Study (Q1 - Q11)

DAEWR gives data (originally from Lew (2007)) from an experiment to determine whether cultured cells respond to two drugs. The study was conducted as a series of six experimental "runs" using a stable cell line plated onto Petri dishes. Each experimental run included three Petri dishes: one treated with drug 1, one treated with drug 2, and one untreated serving as a control. Hence, experiment (Exp) serves as a blocking variable. The data is available from Canvas as "CellData.csv".
See important notes above!

Q1

Provide a visual summary of the data using code similar to what is provided.

```
## 'data.frame':    18 obs. of  3 variables:
##  $ Drug    : chr  "Control" "Control" "Control" "Control" ...
##  $ Exp     : int  1 2 3 4 5 6 1 2 3 4 ...
##  $ Response: int  1147 1273 1216 1046 1108 1265 1169 1323 1276 1240 ...
```



Cell Line RCB1 (Q2 - Q5)

Analyze the data as an RCB design with **fixed blocks** using `lm()`.

Q2

Show the Type 3 ANOVA table.

```
## Anova Table (Type III tests)
##
## Response: Response
##              Sum Sq Df  F value    Pr(>F)
## (Intercept) 2506417  1 266.5723 1.544e-08 ***
## Drug          99122  2   5.2711   0.02734 *
## Exp          134190  5   2.8544   0.07431 .
## Residuals     94024 10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q3

Using the ANOVA table from the previous question, briefly discuss the effectiveness of the blocking for this study.

Response There is significance at a .1 to the Experiment variable which indicates some significant changes between experiments. This can be controlled for as above but may not be entirely necessary as the significance level is only .1

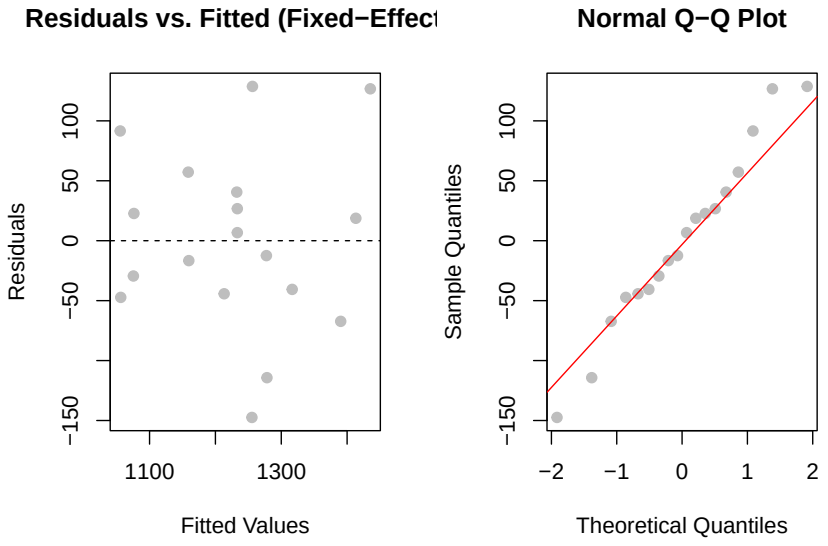
Q4

Run Tukey adjusted pairwise comparisons for Drug and show the results.

```
##  contrast        estimate SE df t.ratio p.value
##  Control - Drug1 -157.833 56 10  -2.819  0.0440
##  Control - Drug2   -0.833 56 10  -0.015  0.9999
##  Drug1 - Drug2    157.000 56 10   2.804  0.0451
##
## Results are averaged over the levels of: Exp
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Q5

Show the residual diagnostic plots (Resids vs Fitted and QQplot of Resids).



Cell Line RCB2 (Q6 - Q9)

Analyze the data as an RCB design with **random blocks** using `lme4::lmer()`.

Q6 (4 pts)

Provide estimates of the block variance and the residual variance.

$\sigma^2_{Block} = \sigma^2_{Exp} = 5811.846$
$\sigma^2_{\epsilon} = \sigma^2_{Resid} = 9402.388$

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Response ~ Drug + (1 | Exp)
##    Data: CellData
##
## REML criterion at convergence: 190.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2329 -0.6221 -0.1807  0.5021  1.6740
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Exp      (Intercept) 5812     76.24
##  Residual             9402     96.97
## Number of obs: 18, groups:  Exp, 6
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 1175.8333    50.3558  23.351
## DrugDrug1    157.8333    55.9833   2.819
## DrugDrug2      0.8333    55.9833   0.015
##
## Correlation of Fixed Effects:
##           (Intr) DrgDr1
## DrugDrug1 -0.556
## DrugDrug2 -0.556  0.500

## Estimated variance for blocks (Exp): 5811.846

## Estimated residual variance: 9402.388
```

Q7

Show the Type 3 ANOVA table. Notes: (1) In the mixed model framework, we can use `anova()` to do this after loading the lmerTest and pbkrtest packages. (2) The F-test for Drug should exactly match your result from Q2.

```
## Analysis of Variance Table
##      npar Sum Sq Mean Sq F value
## Drug    2  99122   49561  5.2711
```

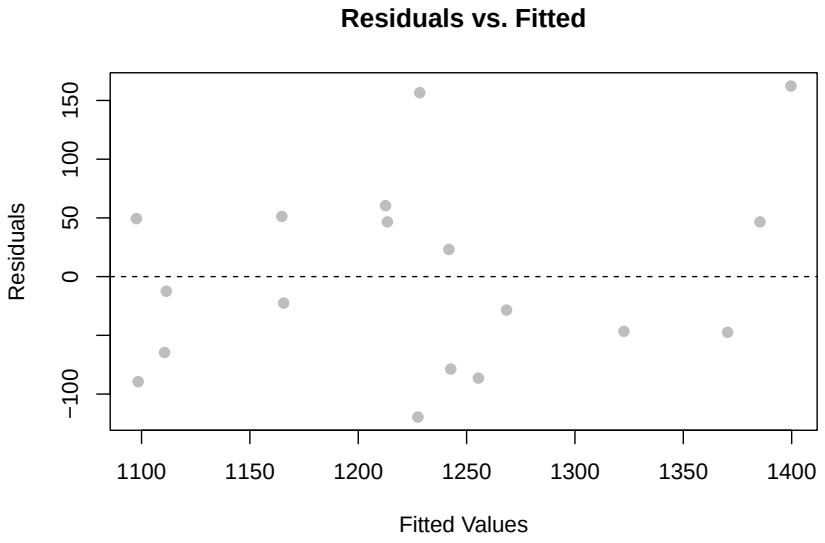
Q8

Run Tukey adjusted pairwise comparisons for Drug and show the results. Note: These results should exactly match the pairwise comparisons from Q4.

```
##  contrast        estimate SE df t.ratio p.value
##  Control - Drug1 -157.833 56 10  -2.819  0.0440
##  Control - Drug2   -0.833 56 10  -0.015  0.9999
##  Drug1 - Drug2    157.000 56 10   2.804  0.0451
##
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Q9

Show the residual diagnostic plot (Resids vs Fitted).



Cell Line CRD (Q10-Q11)

Just for illustration, analyze the data using a one-way model **ignoring blocks**.

Q10

Show the ANOVA table.

```
## Anova Table (Type III tests)
##
## Response: Response
##              Sum Sq Df  F value    Pr(>F)
## (Intercept) 8295504  1 545.2463 3.309e-13 ***
## Drug          99122  2   3.2575   0.06685 .
## Residuals    228214 15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q11

Compare the ANOVA tables from the RCB model (Q2 or Q7) vs the one-way/CRD model (Q10). Briefly explain why we were able to detect differences between Drugs using the RCB model, when we did not detect differences using the one-way/CRD model. Your answer should be based on *specific output*. Hint: You may want to calculate MSResid.

Response the RCB model detects a greater effect with a higher F value and lower residuals than the one-way CRD model. this is largely because the block to block variability which is controleld in questions 2 and 7 is pooled into the error term of question 10

Sleep Study (Q12 - Q16)

DAEWR gives data (originally from le Riche and Csima (1964) from an experiment to evaluate 5 drug treatments (A, B, C, D, E) and their effect on sleep quality in elderly patients. Higher values indicate better sleep quality. A latin square design was used to account for patient-to-patient and week-to-week differences. The data is available from Canvas as "SleepData.csv". A visual summary of the design (and results) is also available from Canvas.
See important notes above!

Q12 (4 pts)

Fit an appropriate fixed effects model using `lm()`. Show the Type 3 ANOVA table.

```
## 'data.frame':    25 obs. of  4 variables:
##  $ Patient: int  1 2 3 4 5 1 2 3 4 5 ...
##  $ Week   : int  1 1 1 1 1 2 2 2 2 2 ...
##  $ Trt    : chr  "B" "D" "E" "A" ...
##  $ Sleep  : num  2.92 2.86 1.97 1.99 2.64 2.43 1.64 2.5 2.39 2.31 ...
```

```
## Anova Table (Type III tests)
##
## Response: Sleep
##             Sum Sq Df  F value    Pr(>F)
## (Intercept) 7.9905  1 167.2382 2.092e-08 ***
## Trt         2.0498  4  10.7255 0.0006202 ***
## Patient     0.6073  4   3.1779 0.0535842 .
## Week        0.3689  4   1.9302 0.1699959
## Residuals   0.5734 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q13

Using the ANOVA table from the previous question, briefly discuss the effectiveness of the latin square design for this study.

Response This design appears effective as we see great signifficance on the treatement variable. There is some significance on Patient and little on the week. We may consider removing week, but the results of the analysis are positive enough that I think we can move forward with this latin square design.

Q14

Run Tukey adjusted pairwise comparisons for Trt and show the results.

```
##  contrast estimate    SE df t.ratio p.value
##  A - B      -0.800 0.138 12  -5.787  0.0007
##  A - C      -0.728 0.138 12  -5.266  0.0015
##  A - D      -0.660 0.138 12  -4.774  0.0034
##  A - E      -0.586 0.138 12  -4.239  0.0083
##  B - C       0.072 0.138 12   0.521  0.9835
##  B - D       0.140 0.138 12   1.013  0.8448
##  B - E       0.214 0.138 12   1.548  0.5537
##  C - D       0.068 0.138 12   0.492  0.9866
##  C - E       0.142 0.138 12   1.027  0.8382
##  D - E       0.074 0.138 12   0.535  0.9818
##
## Results are averaged over the levels of: Patient, Week
## P value adjustment: tukey method for comparing a family of 5 estimates
```

Q15

Considering your results from the previous question, briefly summarize your findings.

Q15A

Discuss comparisons of active treatments (B, C, D, E) vs placebo (A) in context.

Response It appears that all of the treatments are signifficantly different than placebo, but do not differ as much amongst themselves.

Q15B

Discuss comparisons between the active treatments (B, C, D, E) in context.

Response There is no evidence that the treatments are different from eachother. Only that they are different from placebo.

Q16 (4 pts)

Now re-analyze the data as a mixed model using `lme4::lmer()`. Use a model similar to Q12, but decide which terms should be included as fixed and random effects. Show the Type 3 ANOVA table. Note: Due to balance, the F-test for Trt should exactly match your result from Q12.

```
## Type III Analysis of Variance Table with Satterthwaite's method
##     Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## Trt 2.0498 0.51246     4    12  10.726 0.0006203 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix

```r
#Retain this code chunk!!!
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(warning = FALSE)
library(tidyverse)
#Q1

CellData <- read.csv("CellData.csv", stringsAsFactors = FALSE)

str(CellData)

CellData$Exp <- as.factor(CellData$Exp)
CellData$Drug <- as.factor(CellData$Drug)


ggplot(aes(x = Drug, y = Response, group = Exp, color = Exp), data = CellData) +
  geom_point() +
  geom_line()
#Q2
library(car)

lm_fixed <- lm(Response ~ Drug + Exp, data = CellData)

Anova(lm_fixed, type = 3)

#Q4
library(emmeans)

emmeans_fixed <- emmeans(lm_fixed, pairwise ~ Drug, adjust = "tukey")
emmeans_fixed$contrasts
#Q5
par(mfrow = c(1, 2))

plot(lm_fixed$fitted.values, lm_fixed$residuals,
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs. Fitted (Fixed-Effects)",
     pch = 19, col = "grey")
abline(h = 0, lty = 2)

qqnorm(lm_fixed$residuals, main = "Normal Q-Q Plot",
       pch = 19, col = "grey")
qqline(lm_fixed$residuals, col = "red")
#Q6
library(lme4)
lm_random <- lmer(Response ~ Drug + (1 | Exp), data = CellData)

summary(lm_random)

var_comp <- as.data.frame(VarCorr(lm_random))
var_block <- var_comp[var_comp$grp == "Exp", "vcov"]
var_resid <- attr(VarCorr(lm_random), "sc")^2
```

```r
cat("Estimated variance for blocks (Exp):", var_block, "\n")
cat("Estimated residual variance:", var_resid, "\n")

#Q7
library(lmerTest)
anova(lm_random, type = 3)

#Q8
emmeans_random <- emmeans(lm_random, pairwise ~ Drug, adjust = "tukey")
emmeans_random$contrasts
#Q9
plot(fitted(lm_random), residuals(lm_random),
     xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs. Fitted",
     pch = 19, col = "grey")
abline(h = 0, lty = 2)
#Q10
lm_crd <- lm(Response ~ Drug, data = CellData)
Anova(lm_crd, type = 3)
#Q12
SleepData <- read.csv("SleepData.csv", stringsAsFactors = FALSE)

str(SleepData)
SleepData$Patient <- as.factor(SleepData$Patient)
SleepData$Week <- as.factor(SleepData$Week)
SleepData$Trt <- as.factor(SleepData$Trt)

lm_sleep_fixed <- lm(Sleep ~ Trt + Patient + Week, data = SleepData)

Anova(lm_sleep_fixed, type = 3)

#Q14
emmeans_sleep <- emmeans(lm_sleep_fixed, pairwise ~ Trt, adjust = "tukey")
emmeans_sleep$contrasts
#Q16
lm_sleep_mixed <- lmer(Sleep ~ Trt + (1 | Patient) + (1 | Week), data = SleepData)

anova(lm_sleep_mixed, type = 3)
```