Case Study 4 - Group

Alexis, Kamryn, Minh, Addison

For each of the scenarios below, 1) identify the treatments; 2) identify the blocking variables; and 3) describe how you would design this study as a randomized complete block design.

1. A study was conducted to examine the yield of a certain chemical. The treatment factors of interest were acid strength (80% or 90%), time allowed for reaction (15 or 30 minutes), and temperature (50 or 75 degrees Celsius). Of interest was to identify the treatment combination(s) that lead to the highest average yield (in grams).

a. Treatments

Answer: all 6 combinations of acid strength, reaction time, and temperature.

b. Blocking factor

Answer: Lab location would be a blocking factor as there may be lab to lab variation.

c. Design

Answer: RCBD would be used as there is presence of a block and would want to take in account for lab to lab variation. For each lab, samples of the chemical would randomly receive each combination of treatments, such that each lab includes exactly one sample for every treatment combination.

2. Land use along stream banks can affect stream water quality, so farmers often adopt various buffers between their fields and streams. In one study, the DNR is interested in comparing how 3 different buffers affect macro invertebrate populations 100 meters downstream from the buffer. Six different streams have been selected and there is expected to be considerable stream to stream variation in the invertebrate populations.

a. Treatments

Answer: Type of buffer

b. Blocking factor

Answer: Stream

c. Design

Answer: RCBD would be used as there is the presence of a block and would want to take in account for streat to stream variation. For each stream, 3 buffers would be used exactly once on each of the six streams. The buffers would be choosen in random order for each stream.

A team at a research hospital investigated a procedure to create artificial arteries using resin. In this study, resin is obtained from six different batches and a sample of resin from each batch is pressed or extruded through an aperture at one of four different pressures (8500, 8700, 8900, or 9100 PSI) that forms the resin into tubes. The percentage of tubes that meet product specifications for each batch/pressure combination is recorded.

```
#load in data
grafts <- read_csv("grafts.csv")
#convert to factor
grafts <- grafts %>%
  mutate(Batch = factor(Batch),
         Pressure = factor(Pressure))
```

3. Explain why batch should be treated as a blocking factor.

Answer: There may be batch to batch variation due to human error or conditions the batches were made in. To account for the possibility of nuisance we identify batch as a blocking factor. Resin with in each batch would be similar.

4. Before we analyze the data, let's check if the three "core" assumptions are reasonably satisfied.

a. Independence

Answer: We can assume that the resin is independent from eachother, as they are obtained from different batches. eu=ou=resin, resin is not gathered from clusters, and the responses are recorded only once.

b. Equal variance

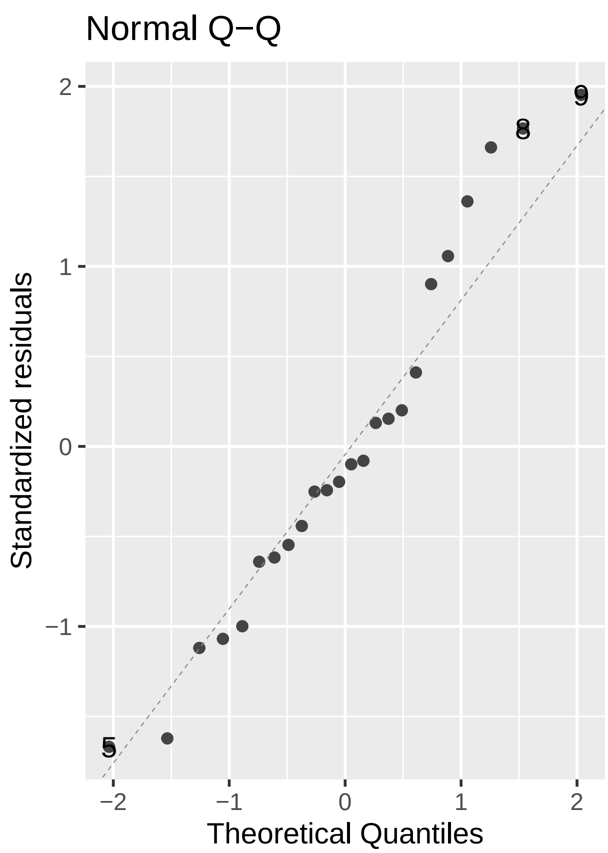
```
favstats(Percentage~Pressure, data = grafts)
```

```
##   Pressure  min     Q1 median     Q3  max     mean       sd n missing
## 1     8500 87.4 89.475  92.10 96.900 98.2 92.81667 4.577081 6       0
## 2     8700 87.0 89.775  91.55 94.150 95.8 91.68333 3.304189 6       0
## 3     8900 85.5 86.650  88.80 90.500 93.4 88.91667 2.966760 6       0
## 4     9100 78.9 83.275  86.50 88.975 90.7 85.76667 4.445072 6       0
```

Answer: Equal Variance assumption is satisfied as 4.6/3.0 < 2.

c. Normality

```
grafts_lm <- lm(Percentage~Pressure+Batch, data = grafts)
autoplot(grafts_lm, which=2)
```



Answer: Normality assumption is satisfied as there are no major deviations at the tails given our sample size of 24.

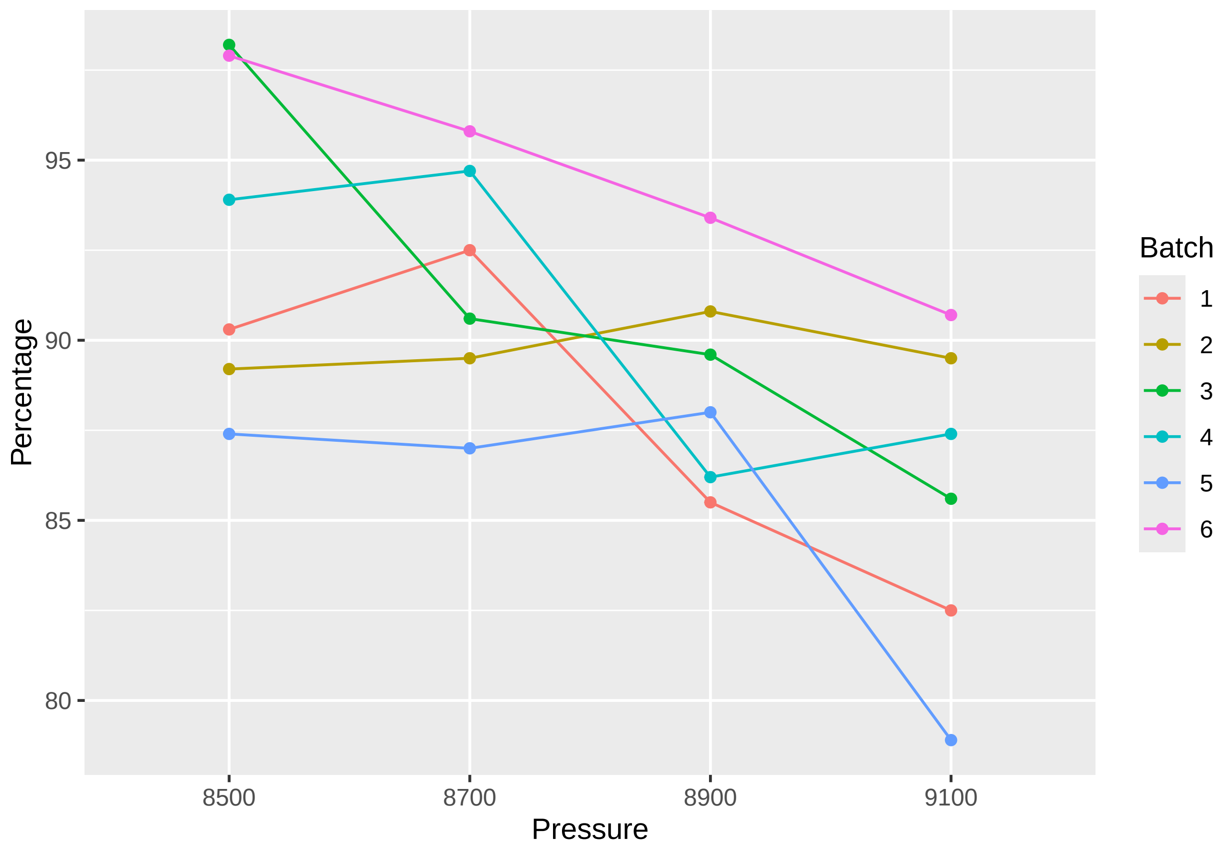
5. Because we only have one replicate per pressure/batch combination, we need to assume that pressure and batch do not interact.

a. Explain what it would mean if pressure and batch do interact.

Answer: If pressure and batch interacted we would not be able to use RCBD as in that model we assume $(\alpha\beta)_{ij} = 0$. If pressure and batch interacted it would mean that the effect of pressure on the response would depend on batch, and it would be unclear what the effect of pressure is. With the data having only one replication per pressure/batch we cannot determine the interaction.

b. Using the appropriate plots, hypothesis tests, etc, explain if it is reasonable to assume that pressure and batch do not interact.

```
#interaction plot
ggplot(data=grafts, aes(x=Pressure, y=Percentage)) +
  geom_line(aes(group=Batch, color=Batch)) +
  geom_point(aes(color=Batch))
```




```r
#tukey
tukey.1df(grafts_lm, data=grafts)
```

```
## $Tukey.SS
## [1] 0.1968562
##
## $Tukey.F
## [1] 0.02512537
##
## $Tukey.p
## [1] 0.8763188
##
## $Devn.SS
## [1] 109.6894
```

Answer: Our interaction plot possibly shows concern with the blocking factor and treatment interacting. To further explore this we will use tukey's test for additivity. H_0: D = 0 vs H_A: D $\neq$ 0 With a p-value of 0.8763 (F=0.025), there is no evidence that pressure and batch interact, thus failing to reject the null hypothesis.

6. Explain whether it was "worth it" to block based on batch using the information provided in the ANOVA table.

```r
anova(grafts_lm)
```

```
## Analysis of Variance Table
##
## Response: Percentage
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Pressure   3 178.17  59.390  8.1071 0.001916 **
## Batch      5 192.25  38.450  5.2487 0.005532 **
## Residuals 15 109.89   7.326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: Based on the interaction plot and tukey's test of additivity we know that the blocking factor of batch would affect the response. In this anova table we see that there is a block effect with a p-value of 0.005 and F = 5.25.

7. Obtain and interpret the efficiency gained in this study by blocking.

```r
grafts_crd <- lm(Percentage~Pressure, data = grafts)
anova(grafts_crd)
```

```
## Analysis of Variance Table
##
## Response: Percentage
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Pressure   3 178.17  59.390  3.9313 0.02345 *
## Residuals 20 302.14  15.107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer 15.1/7.3 = 2.07 If we analyzed the data using CRD we would need 107% more replicates per trt than if we used RCBD.

8. Conduct the appropriate hypothesis test to examine if pressure does affect the percentage of acceptable tubes.

a. Set up the hypotheses.

Answer: H_0: $\alpha_i = 0$ for i = 1, 2, 3, 4 vs H_A at least one $\alpha_i \neq 0$ for i = 1, 2, 3, 4

b. Provide a conclusion for the hypothesis test.

Answer: With a p-value of 0.002 (F = 8.11, df = 3 and 15), there is very strong evidence that pressure affects the percentage of acceptable tubes differs for at least one of the four pressure tubes. Thus we reject the null hypothesis.

9. Based on previous experience, the experimenters believe that as pressure increases, the percentage of acceptable tubes will decrease though they don't have specific comparisons in mind. Using the appropriate post hoc analysis tool(s) investigate whether this is the case. In your answer, make sure to summarize the results of this analysis and to address the experimenters' belief.

```r
graft_means <- emmeans(grafts_lm, ~Pressure)
confint(pairs(graft_means, adjust="tukey"))
```

```
##  contrast                    estimate   SE df lower.CL upper.CL
##  Pressure8500 - Pressure8700     1.13 1.56 15   -3.370     5.64
##  Pressure8500 - Pressure8900     3.90 1.56 15   -0.604     8.40
##  Pressure8500 - Pressure9100     7.05 1.56 15    2.546    11.55
##  Pressure8700 - Pressure8900     2.77 1.56 15   -1.737     7.27
##  Pressure8700 - Pressure9100     5.92 1.56 15    1.413    10.42
##  Pressure8900 - Pressure9100     3.15 1.56 15   -1.354     7.65
##
## Results are averaged over the levels of: Batch
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

```r
confint(contrast(graft_means, list(HvL=c((-1/3), (-1/3), (-1/3), 1))))
```

```
##  contrast estimate   SE df lower.CL upper.CL
##  HvL         -5.37 1.28 15    -8.09    -2.65
##
## Results are averaged over the levels of: Batch
## Confidence level used: 0.95
```

Answer:From this analysis, a higher percentage of tubes are acceptable when placed under pressures of 8500 and 8700 compared to pressure of 9100. Other results are similar to one another, but generally seems to follow the trend that as pressure increases, the percentage of acceptable tubes decreases. For the pressures that differed, with 95% confidence the average percentage of acceptable tubes for the pressure group of 8700 is between 1.4 and 10.4 percent higher than the average percentage of acceptable tubes for the pressure group of 9100. The pressure group of 8500 on average percentage of acceptable tubes is between 2.5 and 11.6 percent higher than pressure group of 9100.