STAT 424 HW4

Fall 2020

Instructions

1. Submit your solution on Canvas as a pdf file before *October 23 at 11:59pm*.
2. Use R for all computation, and include your code.
3. We recommend you start with the template `hw4.Rmd` file on Canvas.

Rubric

- [30 points] One problem below will be graded for,
  - [15 points] Correctness. Are all parts answered correctly?
  - [15 points] Clarity. Are all answers justified? For questions with a coding component, is code provided, and is it readable?
- [20 = 4 * 5 points] Four problems below will be graded on completeness.
- Total: 50 points

```
library("ggplot2")
library("crossdes")
```

```
## Loading required package: AlgDesign

## Loading required package: gtools
```

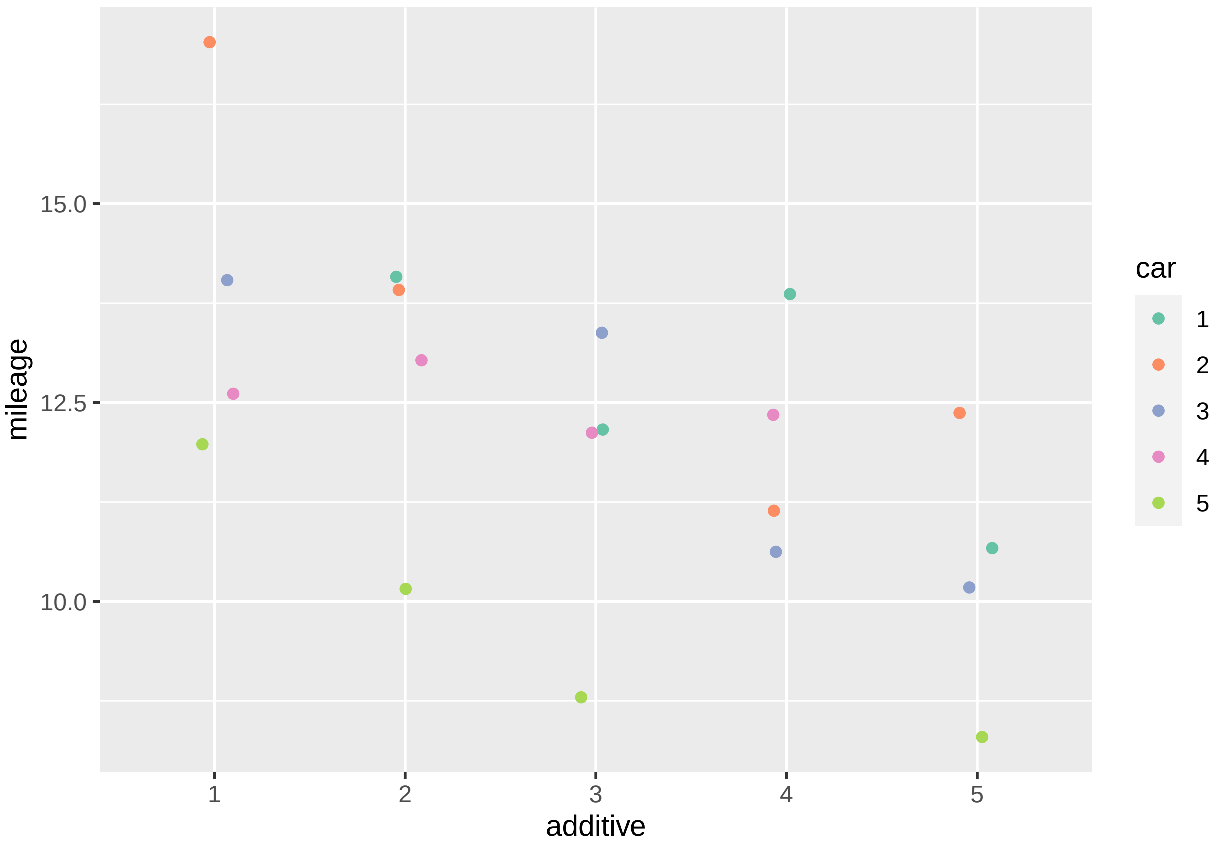
```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

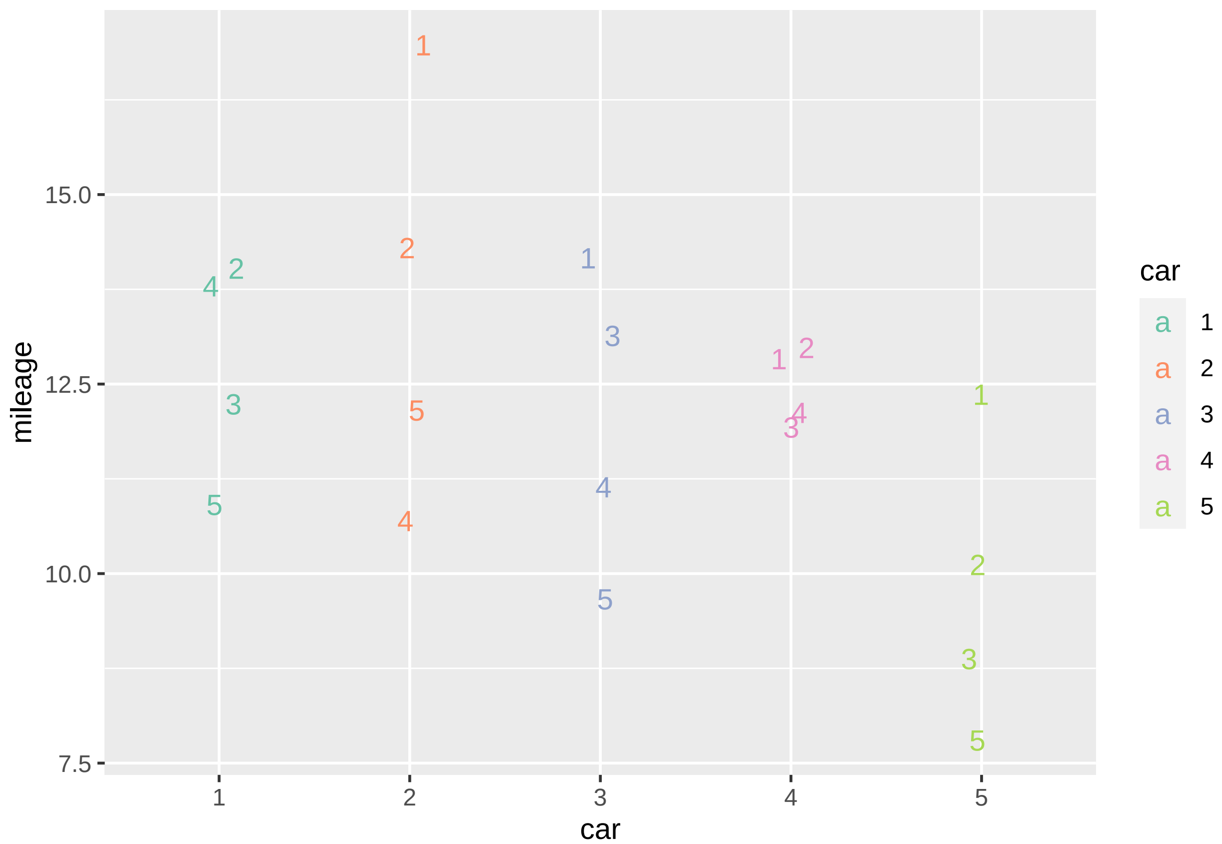
Problems

1. [DAE 4.42] An engineer is studying the mileage performance characteristics of five types of gasoline additives. Ideally, car type would be used as a blocking variable. However, because of time constraints, a complete design is impossible. The following balanced design is run,

```r
mileage <- data.frame(
  additive = rep(1:5, each = 4),
  car = c(2, 3, 4, 5, 1, 2, 4, 5, 1, 3, 4, 5, 1, 2, 3, 4, 1, 2, 3, 5),
  mileage = c(17, 14, 13, 12, 14, 14, 13, 10, 12, 13, 12, 9, 14, 11, 11, 12, 11, 12, 10, 8)
)
mileage$additive=as.factor(mileage$additive)
mileage$car=as.factor(mileage$car)

ggplot(mileage) +
  geom_point(aes(x = additive, y = mileage, col = car), position = position_jitter(w = 0.1)) +
  scale_color_brewer(palette = "Set2")
```




```r
ggplot(mileage) +
  geom_text(aes(x = car, y = mileage, col = car, label = additive), position = position_jitter(w = 0.1))
  scale_color_brewer(palette = "Set2")
```



Analyze data from the experiment (at $\alpha = 0.05$ and draw conclusions).

```
fit1 <- lm(mileage ~ ., data = mileage)
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: mileage
##           Df Sum Sq Mean Sq F value   Pr(>F)
## additive   4 31.300  7.8250  6.9791 0.004744 **
## car        4 36.167  9.0417  8.0642 0.002736 **
## Residuals 11 12.333  1.1212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-values for additive and car are lower than the alpha value of 0.05, they significantly affect the data.

2. [DAE 4.48] An experimenter wishes to compare eight treatments in blocks of four runs. Find a BIBD with 14 blocks and $\lambda = 3$.

```
find.BIB(trt = 8, b = 14, k = 3)
```

```
##      [,1] [,2] [,3]
##  [1,]    2    5    8
```

```
## [2,]    1    3    8
## [3,]    4    6    8
## [4,]    2    5    7
## [5,]    4    7    8
## [6,]    3    5    6
## [7,]    1    3    7
## [8,]    1    2    4
## [9,]    5    6    8
## [10,]   2    3    4
## [11,]   1    5    7
## [12,]   1    2    6
## [13,]   4    6    7
## [14,]   3    4    5
```

3. [DAE 5.7] The yield of a chemical process is being studied. The two most important variables are thought to be pressure and temperature. Three levels of each factor are selected, and a factorial experiment with two replicates is performed. The yield data are as follows.

```
chemical <- data.frame(
  temperature = rep(c(150, 160, 170), each = 6),
  pressure = rep(c(200, 215, 230), 6),
  yield = c(90.4, 90.7, 90.2, 90.2, 90.6, 90.4,
            90.1, 90.5, 89.9, 90.3, 90.6, 90.1,
            90.5, 90.8, 90.4, 90.7, 90.9, 90.1)) %>%
  mutate(
    temperature = as.factor(temperature),
    pressure = as.factor(pressure)
  )
```

a. Analyze the data and draw conclusions. Use $\alpha = 0.05$.

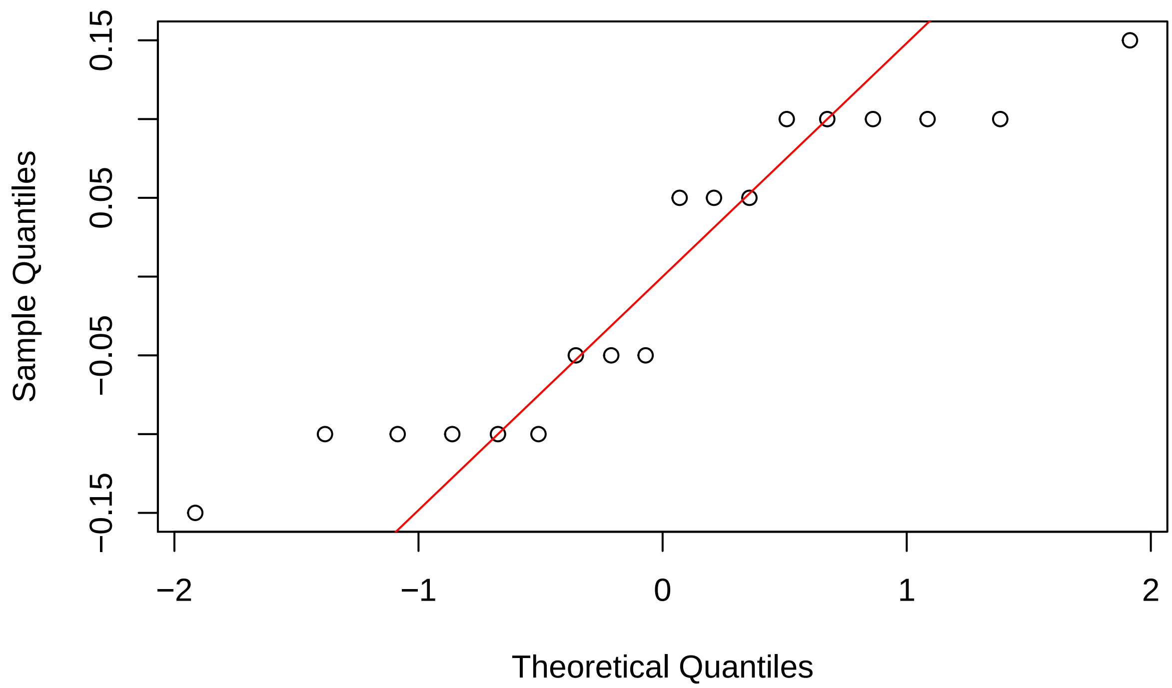
```
fit3 <- lm(yield ~ temperature * pressure, data = chemical)
anova(fit3)
```

```
## Analysis of Variance Table
##
## Response: yield
##                      Df  Sum Sq Mean Sq F value    Pr(>F)
## temperature           2 0.30111 0.15056  8.4687 0.0085392 **
## pressure              2 0.76778 0.38389 21.5937 0.0003673 ***
## temperature:pressure  4 0.06889 0.01722  0.9687 0.4700058
## Residuals             9 0.16000 0.01778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p=value for the temperature and pressure is lower than the alpha value of 0.05, they significantly affects the data. The interaction is not lower, so it does not significantly affect the data.

b. Prepare appropriate residual plots and comment on the model's adequacy.

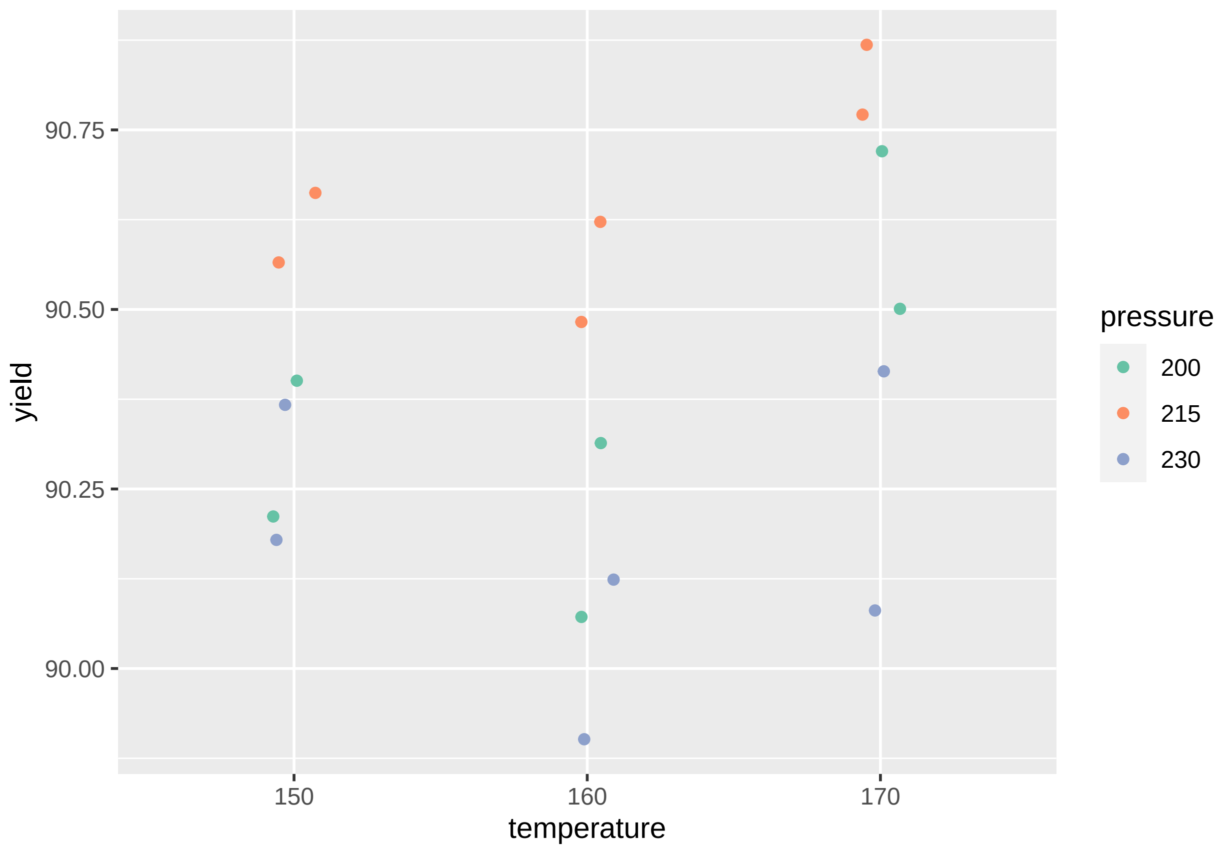
```r
qqnorm(resid(fit3), main = "")
qqline(resid(fit3), col = "red")
```



The model is slightly skewed right.

c. Under what conditions would you operate this process?

```r
ggplot(chemical) +
  geom_point(aes(x = temperature, y = yield, col = pressure), position = position_jitter(w = 0.1)) +
  scale_color_brewer(palette = "Set2")
```



Highest yield is when pressure = 215 and temperature = 170.

4. [DAE 5.26] An article in the IEEE Transactions on Electron Devices describes a study on polysilicon doping. The experiment shown below is a variation of their study. The response variable is base current.

```
silicon <- data.frame(
  ions = rep(1:2, each = 6),
  temperature = rep(c(900, 950, 1000), 4),
  current = c(4.6, 10.15, 11.01, 4.40, 10.2, 10.58,
              3.20, 9.38, 10.81, 3.50, 10.02, 10.6)) %>%
  mutate(
    ions = as.factor(ions),
  )
```

a. Is there evidence (at $\alpha = 0.05$) indicating that either polysilicon doping level or anneal temperature affects base current?

```
fit4 = lm(current ~ ions + as.factor(temperature), data = silicon)
anova(fit4)
```

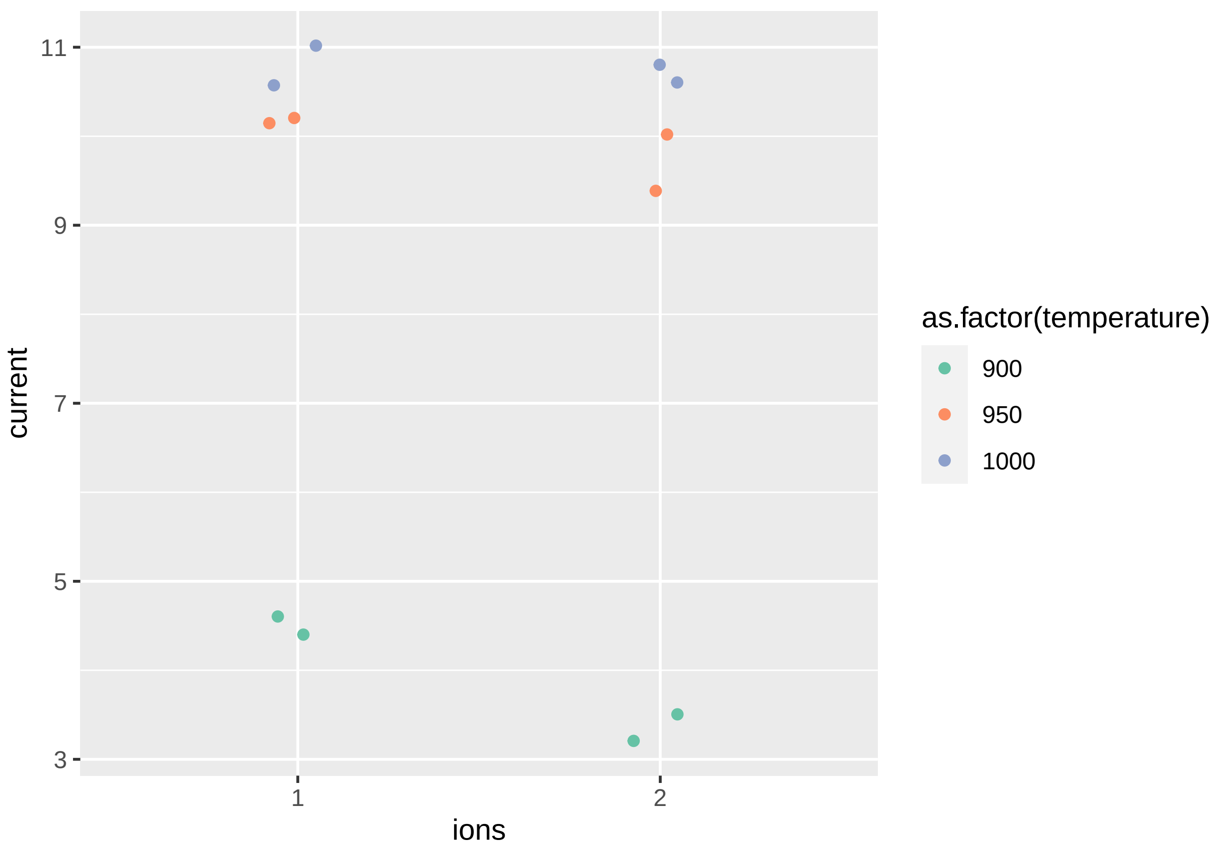
```
## Analysis of Variance Table
##
## Response: current
##                         Df  Sum Sq Mean Sq  F value   Pr(>F)
```

```
## ions                      1   0.980   0.980   8.1585 0.02127 *
## as.factor(temperature)  2 111.188  55.594 462.6244 5.4e-09 ***
## Residuals                 8   0.961   0.120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

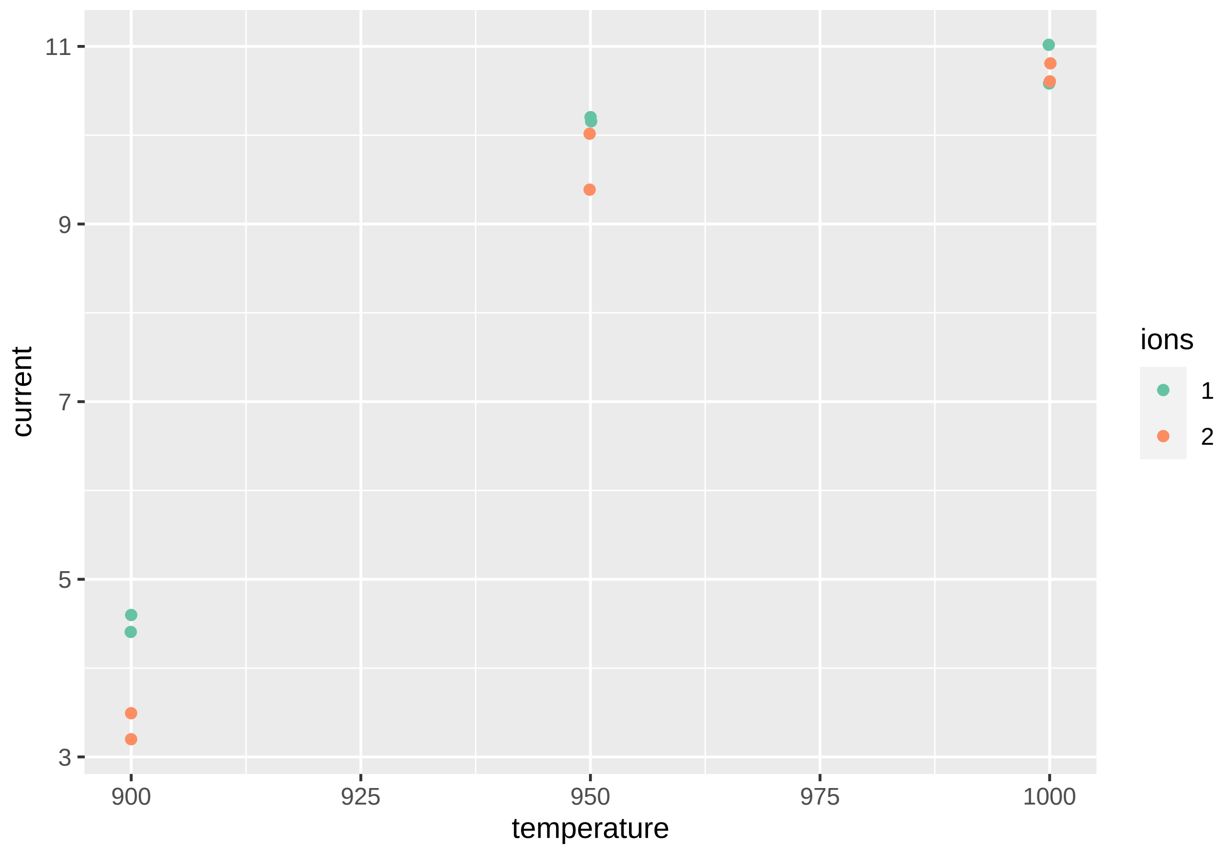
They do affect base current, since the p-value of 0.02127 and 5.4e-09 are lower than the alpha value of 0.05.

b. Prepare graphical displays to assist in interpreting this experiment.

```
ggplot(silicon) +
  geom_point(aes(x = ions, y = current, col = as.factor(temperature)), position = position_jitter(w = 0
  scale_color_brewer(palette = "Set2")
```

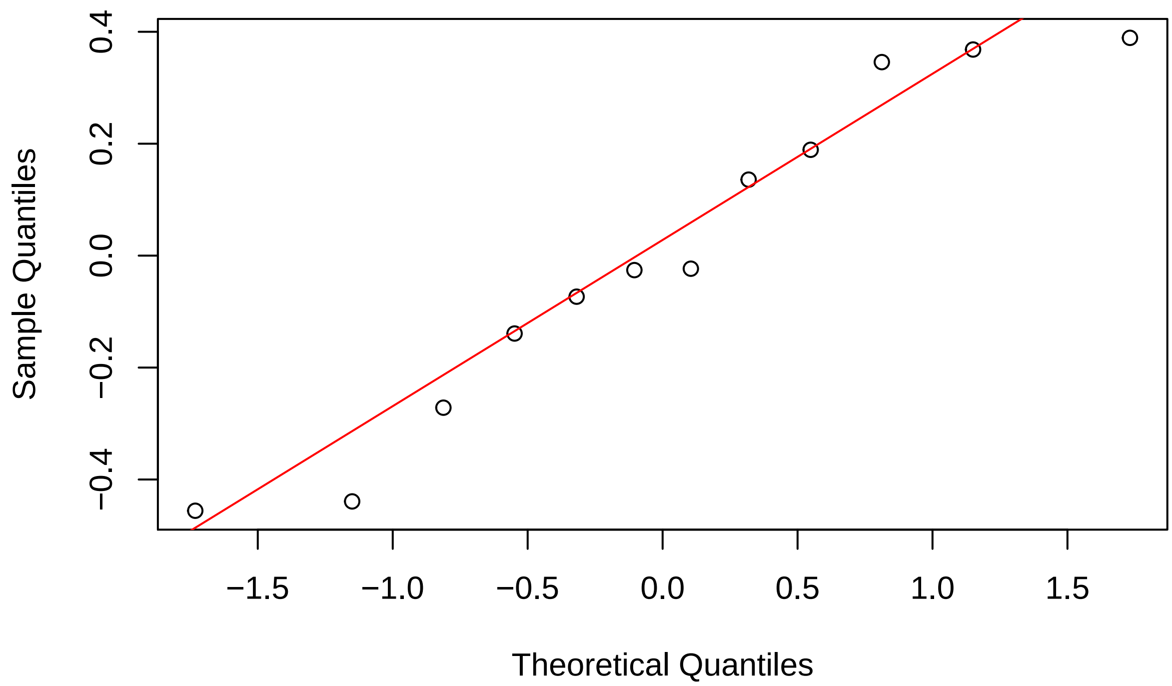



```
ggplot(silicon) +
  geom_point(aes(x = temperature, y = current, col = ions), position = position_jitter(w = 0.1)) +
  scale_color_brewer(palette = "Set2")
```



c. Analyze the residuals and comment on model adequacy.

```
qqnorm(resid(fit4), main = "")
qqline(resid(fit4), col = "red")
```



The model is slightly skewed left.

d. Is the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$$

supported by this experiment (x_1 is doping level and x_2 is temperature)? Estimate the parameters in this model and plot the response surface.

```
silicon.2 = silicon
silicon.2$ions = as.numeric(silicon.2$ions)

fit5 = lm(current ~ ions * temperature + I(temperature^2), data = silicon.2)
sum.fit5=summary(fit5)
sum.fit5
```

```
##
## Call:
## lm(formula = current ~ ions * temperature + I(temperature^2),
##     data = silicon.2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27167 -0.14042 -0.04833  0.12458  0.36833
##
```

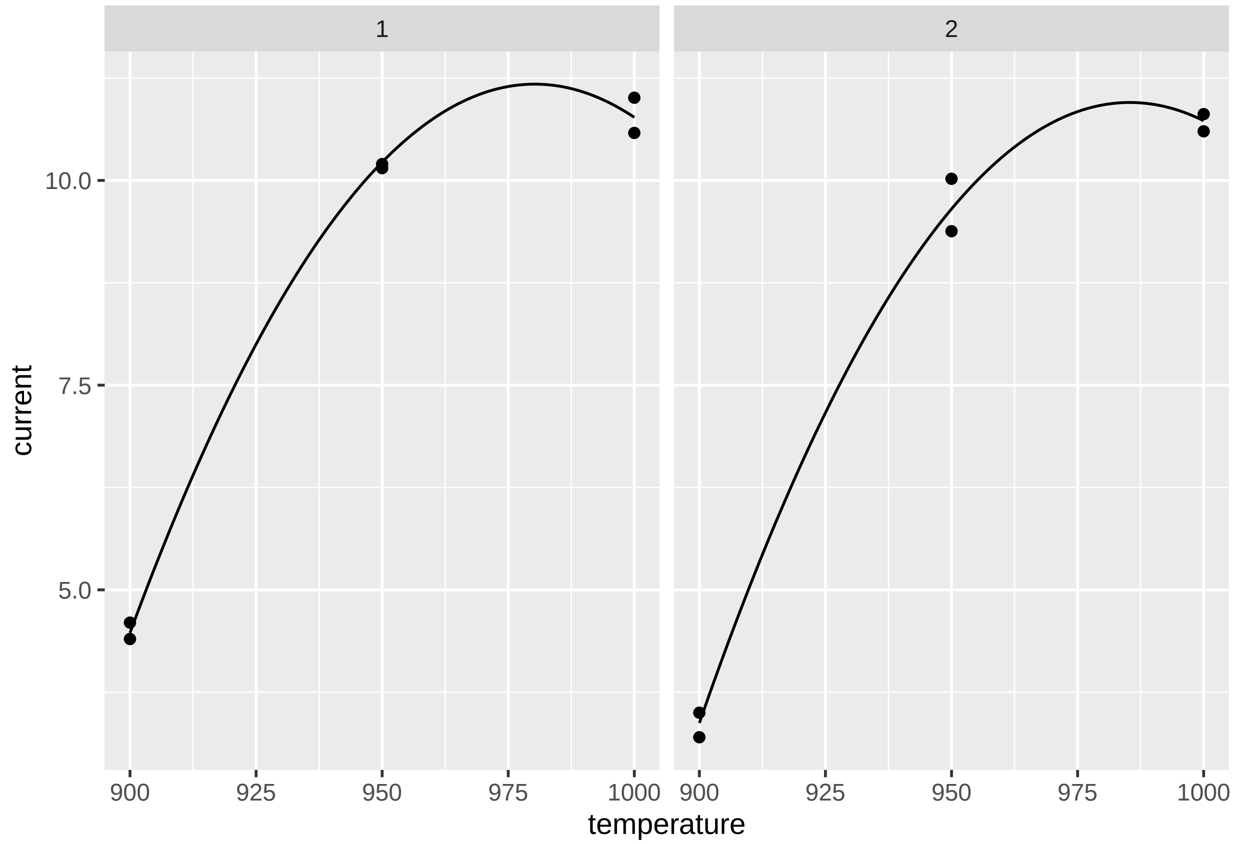
```
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)    
## (Intercept)      -9.775e+02  5.296e+01 -18.457 3.40e-07 ***
## ions             -1.064e+01  3.213e+00  -3.312   0.0129 *  
## temperature       2.028e+00  1.113e-01  18.221 3.71e-07 ***
## I(temperature^2) -1.040e-03  5.852e-05 -17.771 4.41e-07 ***
## ions:temperature  1.060e-02  3.379e-03   3.137   0.0164 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.2389 on 7 degrees of freedom
## Multiple R-squared:  0.9965, Adjusted R-squared:  0.9944 
## F-statistic: 493.7 on 4 and 7 DF,  p-value: 1.175e-08
```

Since the p-vale of 1.175e-08 is lower than the alpha value of 0.05, the model is adequate.

```r
intercept = sum.fit5$coefficients[1,1]
beta1 = sum.fit5$coefficients[2,1]
beta2 = sum.fit5$coefficients[3,1]
beta22 = sum.fit5$coefficients[4,1]
beta12 = sum.fit5$coefficients[5,1]
list(intercept = intercept, beta1 = beta1, beta2 = beta2, beta22 = beta22, beta12 = beta12)
```

```
## $intercept
## [1] -977.5375
## 
## $beta1
## [1] -10.64167
## 
## $beta2
## [1] 2.02835
## 
## $beta22
## [1] -0.00104
## 
## $beta12
## [1] 0.0106
```

```r
silicon.new <- expand.grid(
  ions = unique(silicon.2$ions),
  temperature = seq(900, 1000, by = 1)
  )
silicon.new$current<- predict(fit5, silicon.new)

ggplot() +
  geom_point(data = silicon, aes(x = temperature, y = current)) +
  geom_line(data = silicon.new, aes(x = temperature, y = current)) +
  facet_wrap(~ ions)
```



5. Discussion of interaction terms.

a. In your own words, explain what it means for there to be an interaction between two factors.

It means that the outcome of one factor in a test is dependent on the outcome of another factor in a given test.

b. In your own words, explain what it means for there to be an interaction between three factors.

It means that the outcome of one factor in a test is dependent on the outcome of two separate factors that are dependent on each other.

c. In real experiments, what do you expect to be more common, two-way interactions or three-way interactions? Why? Note: you can argue either way, as long as it's well-justified. Contemplating this question will prepare you for fractional factorial designs.

I believe that three way interactions are more common. Each variable will take in all other factors and respond a certain way. For example, let's say someone is testing the effect of hydration, athletic performance, and nutrition on likelihood for obesity. If someone drinks and eats right, they will probably be physically fit because they take care of their bodies. Any two combinations of the dependent variables will affect the third because they can influence how a person takes care of their body.

Feedback

a. How much time did you spend on this homework?

A total of 5 hours.

b. Which problem did you find most valuable?

Problem 4d. I liked learning about the different types of Rcode that can be used to influence a model.

References

[DAE] Montgomery, Douglas C. Design and analysis of experiments. John wiley & sons, 2017.