Stat 509: Statistics for Engineers

Homework Assignment 10

Make sure to include the R code and graphs used in your analyses.
You may assume $\alpha = 0.05$ for any significance tests.

1. You have data for the lifetime in years of a certain type of wire, the thickness in mm of the coating on the wire, and the temperature in Celsius of the wire's surroundings. You wish to create a linear regression model for the lifetime of a wire based upon the thickness of its coating and the surrounding temperature. Read the *wire* data into R in order to answer the following questions:

   (a) Create scatterplots for the response against each predictor. Do there appear to be any relationships?

plot(coat, life, xlab = "Thickness (mm)", ylab = "Lifetime (years)", main = "Lifetime vs Thickness")

plot(temp, life, xlab = "Temperature (Celsius)", ylab = "Lifetime (years)", main = "Lifetime vs Temperature")

Coat & Life seems to have a relationship.

Temperature & Life doesn't.

   (b) Fit the following linear regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \epsilon,$$

   where Y is the lifetime of the wire, x_1 is the wire's thickness, and x_2 is the temperature. Can any of the predictors can be dropped from the model? If so, refit the model accordingly. Be thorough in the description of what leads to your choice of model and include supporting evidence.

model = lm(life ~ coat + temp)
summary(model)

12.68 + 2.33*coat + 0.0045*temp

2. You are charge of ensuring the structural integrity of a new building. You have collected data on the structural strength of 40 buildings, which were built using one of two types of alloys and using one of two designs. Read the *structure* data into R in order to answer the following questions:

   Fit the following linear regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

   ~~If the interaction is **not** needed:~~

   ~~(a) Fit the model without the interaction.~~

   ~~(b) Provide evidence that the model fits properly and that the model assumptions are satisfied.~~

   ~~(c) Interpret the estimated regression coefficients.~~

   ~~(d) Do we have evidence that there is an optimal way to build the new structure?~~

   If the interaction **is** needed:

   (a) Explain what the interaction means in the context of the problem.

```
model <- lm(strength ~ alloy + design + alloy * design)
summary(model)
```

The interaction is between alloy and design. The interaction shows how different types of alloys affect the design.

(b) Provide evidence that the model fits properly and that the model assumptions are satisfied.

```
plot(model, which = 1)
```
In the Residuals vs Fitted graph, the line is nearly linear and the points are evenly distributed.
```
plot(model, which = 2)
```
In the Q-Q Residuals graph, is linear and the points are seen on both sides of the line.

Therefore, the model fits properly and the model assumptions are satisfied.

(c) Perform a one-way ANOVA analysis, ignoring the factorial treatment structure.

```
anova(lm(strength ~ alloy, data = data2))
```
Pr(>F) = 0.001058

```
anova(lm(strength ~ design, data = data2))
```
Pr(>F) = 6.818e-13

(d) Provide evidence that the one-way ANOVA assumptions are satisfied.

Based on the P values from c, we can assume that all assumptions are satisfied because they each don't exceed 0.05.

(e) If there appears to be a difference in the mean strength among the treatments perform a Tukey procedure at the 95% family-wise confidence level and explain the information it provides.

```
TukeyHSD(aov(strength ~ alloy, data = data2))
```

```
TukeyHSD(aov(strength ~ design, data = data2))
```

The information provided by this test shows that the alloy and design have a significant impact on the structure of the building.

3. You have collected data for the time in minutes that it takes for a compound to solidify. You have also collected data on the temperature (0=low, 1=high) and the humidity (0=low, 1=high). Read the *solidify* data into R in order to answer the following questions:

   Fit the following linear regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

   ~~If the interaction is **not** needed:~~

   ~~(a) Fit the model without the interaction.~~

   ~~(b) Provide evidence that the model fits properly and that the model assumptions are satisfied.~~

   ~~(c) Interpret the estimated regression coefficients.~~

   ~~(d) Do we have evidence that one of ways to solidify the compound is faster than the others?~~

   If the interaction **is** needed:

   (a) Explain what the interaction means in the context of the problem.

```
model = lm(time ~ temp + humid + temp * humid, data = data3)
summary(model)
```

The interaction shows that the temperature and humidity affect the time.

   (b) Provide evidence that the model fits properly and that the model assumptions are satisfied.

All p values are lower than 0.05. Therefore, the model assumptions are satisfied.

   (c) Perform a one-way ANOVA analysis, ignoring the factorial treatment structure.

```
anova(lm(time ~ temp, data = data3))
P value = 0.0006722

anova(lm(time ~ humid, data = data3))
P value = 7.368e-09
```

   (d) Provide evidence that the one-way ANOVA assumptions are satisfied.

Both p values are less than 0.05. Therefore, the assumptions are satisfied.

   (e) If there appears to be a difference in the mean solidification time among the treatments perform a Tukey procedure at the 95% family-wise confidence level and explain the information it provides.

```
TukeyHSD(aov(time ~ temp, data = data3))

TukeyHSD(aov(time ~ humid, data = data3))
```

These tests show that the temperature and humidity interactions have a significant impact on the time.