STAT 308 – Homework 7

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you will not be given full credit. Please upload your assignment by Thursday, November 3, 11:59 pm in a pdf file to Sakai.

1. For each of the following studies/experiments, determine what the numerator and denominator degrees of freedom for performing an F test for differences in the group means will be. In other words, what are the degrees of freedom for the model and for the error?

(a). We wish to determine if the average GPA of Loyola students is different between the four different grades (freshman, sophomore, junior, senior). We sample a total of 50 students in each grade, record their GPA, and perform an F test for differences in the mean GPA between the grades.

$df_{num} = 4 - 1 = 3$, $df_{den} = 50 * 4 - 4 = 196$

(b). In an experiment, 50 randomly selected overweight adults in the same geographic area are randomly assigned either a weight loss drug or a placebo. After 6 weeks, we record the amount of weight each adult lost and perform and F test for a difference in the average weight loss between the adults who received the drug and those who received the placebo.

$df_{num} = 2 - 1 = 1$, $df_{den} = 50 - 2 = 48$

(c). The United States is divided into 5 unique regions: northeast, southeast, midwest, southwest, and west. We randomly sample 1000 households in each region (5000 total households), and wish to perform an F test to determine if the average household income is different between the 5 regions.

$df_{num} = 5 - 1 = 4$, $df_{den} = 1000 - 5 = 995$

(d). In a school experiment, students apply three different types of fertilizer to 12 total potted plants. After 1 week, the students measure the growth of the plant and determine if there is a difference between the mean plant growth between the three different types of fertilizer.

$df_{num} = 3 - 1 = 2$, $df_{den} = 12 - 3 = 9$

2. Consider the dataset `ToothGrowth` available in base `R`, a dataset based on a study on the length of odontoblasts (cells repsonsible for tooth growth) in 60 guinea pigs. Each guinea pig received either 0.5, 1, or 2 mg/day of Vitamin C by either orange juice (OJ) or absorbic acid (VC). The dataset contains the following three variables:

- `len`: The numeric length of the tooth
- `supp`: The Vitamin C supplement type (either VC or OJ)
- `dose`: The dosage in mg/day

Use this dataset to answer the following questions:

a. Fit a linear model for length of the teeth by dosage of the amount of Vitamin C the guinea pigs are receiving. Perform a formal hypothesis test to determine if there is a linear relationship between Vitamin C dosage and tooth length. Assume $\alpha = 0.05$.

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

```
mod <- lm(len ~ dose,ToothGrowth)
f <- summary(mod)$f
f["value"]
```

```
##    value
## 105.0648
```

```
pf(f["value"],f["numdf"],f["dendf"],lower.tail=FALSE)
```

```
##        value
## 1.232698e-14
```

$f = 105.0648$, $p = 1.233e-14$. Because $p < \alpha$, we reject H_0 and say that there is a significant linear relationship between vitamin C dosage and guinea pig tooth length.

b. Now, suppose the experimentors wish to know if adding the Vitamin C supplement to our linear model significantly improves of the predictive ability of tooth length.

(i). Provide least squares regression lines from this linear model for each of the two Vitamin C supplement types.

```
mod_full <- lm(len ~ dose + supp,ToothGrowth)
```

$\hat{Len}_{OJ} = 9.273 + 9.764Dose$ $\hat{Len}_{VC} = 5.573 + 9.764Dose$

(ii). What is the baseline value for supplement type that `R` creates?

The baseline is "Orange Juice''.

(iii). Interpret the value of the parameter associated with the indicator variable added to our model in the context of the problem. (Hint: Think of what the interpretation of the intercepts for the linear models in (i) are.)

Holding dosage constant, we expect the tooth length of guinea pigs who received the ascorbic acid to be 3.7 mm shorter than the guinea pigs who received orange juice.

(iv). Perform an formal F test to determine if adding Vitamin C supplement to our linear model significantly improves of the predictive ability of tooth length. Make sure it is clear which parameter in the model you are testing for. Assume $\alpha = 0.05$.

Reduced Model: $\hat{Len} = \beta_0 + \beta_1 Dose$

Full Model: $\hat{Len} = \beta_0 + \beta_1 Dose + \beta_2 Supp_{VC}$

$$H_0 : \beta_2 = 0$$
$$H_a : \beta_2 \neq 0$$

```
mod_red <- lm(len ~ dose,ToothGrowth)
anova(mod_red,mod_full)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ dose
## Model 2: len ~ dose + supp
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     58 1227.9
## 2     57 1022.6  1    205.35 11.447 0.001301 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$F = 11.447$, $p = 0.001301$. We reject H_0 and can conclude that including the supplement type in our linear model that already includes dosage amount signficantly improves the predictive ability of guinea pig tooth length.

c. Now, suppose we believe that the linear relationship between dosage and tooth length is different for the two different supplement types. In other words, we wish to add an interaction term for dosage and supplement type to our model.

(i). Provide least squares regression lines from this linear model for each of the two Vitamin C supplement types.

```
mod_full <- lm(len ~ dose*supp,ToothGrowth)
```

$\hat{len}_{OJ} = 11.550 + 7.811 Dose$

$\hat{len}_{VC} = 3.295 + 11.715 Dose$

(ii). Interpret the value of the parameter associated with the interaction term in our model in the context of the problem. (Hint: Think of what the interpretation of the slopes for the linear models in (i) are.)

When dosage increases by 1 mg/day, we expect tooth length to increase by 3.904 mm more for the guinea pigs who receive the ascorbic acid than the those who received orange juice.

(iii). Perform an formal F test to determine if adding the linear relationship between dosage and tooth length is different between the two different supplement types. Make sure it is clear which parameter in the model you are testing for. Assume $\alpha = 0.05$.

Reduced Model: $\hat{Len} = \beta_0 + \beta_1 Dose + \beta_2 Supp_{VC}$

Full Model: $\hat{Len} = \beta_0 + \beta_1 Dose + \beta_2 Supp_{VC} + \beta_3 Dose \times Supp_{VC}$

$$H_0 : \beta_3 = 0$$
$$H_a : \beta_3 \neq 0$$

```
mod_red <- lm(len ~ dose + supp,ToothGrowth)
anova(mod_red,mod_full)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ dose + supp
## Model 2: len ~ dose * supp
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1     57 1022.56
## 2     56  933.63  1     88.92 5.3335 0.02463 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$F = 5.334$, $p = 0.0246$. We reject H_0 and can conclude that the linear relationship between dosage and guinea pig tooth length is different between the two different Vitamin C supplement types.