2028: Basic Statistical Methods
Homework 7

This homework is due Wednesday, Nov. 20th in class **BEFORE class starts**. Late papers will not be accepted. Please do not turn in any papers to any mailbox.

- Please remember to staple if you turn in more than one page.
- Please make sure to **SHOW ALL WORK** in order to receive full credit.

1 Find the missing values in the analysis of variance table below.

Source	Degrees of freedom	Sum of Squares	Mean Squares	F-statistics	p-value
Treatments	?	?	111.4	?	?
Error	23	461.9	?		
Total	28	?			

2 A balanced experimental design has sample size of $n = 12$ observations at each of $k = 7$ factor levels. The total sum of squares is $SST = 133.18$, and the sample averages are $\bar{x}_{1\cdot} = 7.75$, $\bar{x}_{2\cdot} = 8.41$, $\bar{x}_{3\cdot} = 8.07$, $\bar{x}_{4\cdot} = 8.30$, $\bar{x}_{5\cdot} = 8.17$, $\bar{x}_{6\cdot} = 8.81$, and $\bar{x}_{7\cdot} = 8.32$. Compute the ANOVA table. (*Hint*: Calculate $SSTr$ and then subtract it from SST to obtain SSE).

3 Textbook 13-2 page 530. The data is called data132.csv (data 6).

4 **Hot dogs**

People who are concerned about their health may prefer hot dogs that are low in salt and calories. The "Hot dogs" datafile contains data on the sodium and calories contained in each of 54 major hot dog brands based on the results of a laboratory analysis of calories and sodium content. The hot dogs are classified by type: beef, poultry, and meat (mostly pork and beef, but up to 15% poultry meat).

Variable Names:

1. Type: Type of hotdog (beef, meat, or poultry)
2. Calories: Calories per hot dog
3. Sodium: Milligrams of sodium per hot dog

Reference: Moore, David S., and George P. McCabe (1989). Introduction to the Practice of Statistics. Original source: Consumer Reports, June 1986, pp. 366-367.

Getting the Data: The file name is called *hotdogs.txt* (data 7). Once you have saved the data file in the working directory, read the data in R using the command

```
data = read.table("hotdogs.txt",header=TRUE)
```

In the analysis below, we are only interested in difference of mean calories for each type of hot dog.

```
calories = data[,2]
type = data[,1]
```

Question 1. Conduct exploratory analysis of the data: What does your exploratory analysis suggest about how the number of calories is related to the type of hotdogs?

The R command for constructing boxplots of a variable (score) categorized by a second variable (condition) is

```
boxplot(V1~V2)
```

Question 2. Perform an ANOVA of the score with respect to condition.

To fit the One-Way ANOVA model to the data with R, you will apply the function `aov`. The input is the formula $V1\tilde{}V2$ where $V1$ is the numerical variable categorized according to $V2$. The R commands are

```
model = aov(calories~type)
```

To output of the ANOVA model is

```
summary(model)
model.tables(model,type = "means")
```

Based on this output, answer the following questions:

(i) What is the *sum of squares total*, *sum of squares of treatments* and *sum of squared error*?

(ii) What are the model parameters? Define clearly what each of these parameters represents (in context).

(iii) What are the estimates of these parameters? Use the output from ANOVA to obtain these estimates.

Question 3. Do the assumption of ANOVA hold for this example?

In order to obtain the residual plots for diagnostics run the R command

```
par(mfrow=c(2,2))
plot(model)
```

Enumerate each assumption and show whether you think it holds or not and why.

Question 4. For this example, perform the ANOVA F Test.

(i) State the appropriate H_0 and H_a.

(ii)

• What is the F test statistic?

• Now write it in a way that shows how F is obtained from the ANOVA table. In other words, write the following:

$$F = \frac{SSTr/(k-1)}{SSE/(N-k)} == \frac{\ /\ }{\ /\ }$$

• Fill in the blanks: Under H_0 (i.e., when the null hypothesis is true), the F test statistic has an

$F(\ ,\ )$ distribution.

(This is the distribution under which the p-value is calculated).

(iii) What is the p-value of the ANOVA F test?

(iv) State you conclusions in context.

Question 5. Now, we would like to learn more about the nature of the effect of the type of the hot dogs. In other words, we would like to find, which of the μ's are not equal. To this end we will use the procedure which tests each pair of means (μ_i, μ_j) to determine whether they are equal - the "multiple (paired) comparisons":

$$H_0 : \mu_i = \mu_j \ H_A : \mu_i \neq \mu_j$$

while maintaining the overall (family) error rate at 5%.

For each pair (μ_i, μ_j) the procedure provides a confidence interval for $\mu_i - \mu_j$, so we get a family of confidence intervals. Testing $H_0 : \mu_i = \mu_j$ $H_A : \mu_i \neq \mu_j$ is equivalent to checking whether 0 falls inside the confidence interval for $\mu_i - \mu_j$ or not. In other words, we check whether 0 is a plausible value for the difference $\mu_i - \mu_j$ or not. If it is (0 in the confidence interval), than we have no evidence to reject that $\mu_i = \mu_j$.

To obtain the pairwise confidence intervals, use the following R command

```
TukeyHSD(model)
```

Write down ONLY the pair(s) of means (μ_i, μ_j) which, based on the output, are significantly different. Summarize your findings in context of the problem.