**BF[1] – Using Software**

**Type in your score here 🡪 \_\_41\_\_ out of 41 points possible**

1. (5 points) Ponder/Reflect Exercise – Reflect on what you have learned from this portion of the class. Examples of what you can do are: a brief outline of material covered, insights you gained from class or personal study, or items you feel that you need to follow up or work on. (3-5 sentences)

Some insights I have gained is that visualizations can help determine what kind of tests to run on data. Also, that knowing how to set up an ANOVA table by hand it is easier to understand when doing it by software such as R. Software can be very helpful when doing statistical tests.

2. Kudzu is a plant that was imported to the United States from Japan and now covers over seven million acres in the South. The plant contains chemicals called isoflavones that have been shown to have beneficial effects on bones. One study used three groups of rats to compare a control group with rats that were fed either a low dose or a high dose of isoflavones from kudzu. One of the outcomes examined was the bone mineral density in the femur (in grams per square centimeter). You would like to test if the mean bone mineral density is different for the three different groups. Use α = 0.05 level of significance.

*Check Requirements and Descriptive Statistics*

(a) (1 point) What descriptive statistics would you use to describe the data (both numerical and graphical)? Please show the descriptive statistics.

(b) (1 point) Check the Requirement that the residuals are normally distributed.

(c) (1 point) Does the Requirement of equal standard deviations or variances hold? Why?

It appears that the Requirement holds in this case since the largest standard deviation divided by the smallest standard deviation is less than 2.

*Inferential Statistics*

(d) (2 points) Write the appropriate null and alternative hypotheses.

Ho: µ1= µ2 = µ3 Ha: At least one mean is different than the others

(e) (1 point) Report the sample test statistic.

7.718

(f) (1 point) State the P-value.

0.0014

(g) (1 point) What decision do you make based on the P-value and level of significance (α)?

Since the p-value is less then the level of significance we would reject the null hypothesis.

(h) (1 point) State your conclusions in “layman’s terms”.

We have sufficient evidence at least one of the means of the groups is different then the 3 other groups.

3. USING R, do a complete analysis of variance comparing mean survival times for the cancer types. (For this exercise, do NOT log transform the data-just use the raw values.) For full credit, show the appropriate output.

(a) (2 points) Calculate and list the mean survival time in days for each cancer type.

(b) (2 points) Check the requirements that the residuals are normally distributed and equal variances.

(c) (6 points) Test the hypothesis

H0: All cancer types have the same mean survival time.

1. State the null and alternative hypotheses,

**Ho: All of the population means are equal**

**Ha: At least one of the population means is different**

1. Give the ANOVA table
2. Give and Interpret the F statistic

4.793

1. **The ratio of between group variation vs chance error is 4.793. It appears that we might have a difference for at least one of the groups**
2. Give and interpret the p-value

0.001

**The probability of calculating the test statistic we got or more extreme assuming the null hypothesis is true is 0.001.**

1. State the degrees of freedom for the F statistic,

**Degrees of freedom for numerator = 5 Degrees of Freedom for Denominator = 57**

1. Make a conclusion.

**The F-statistic is large and the p-value is small, we would reject the null hypothesis. We would then say that at least one of the means is different**

4. USING R, do a complete analysis of variance comparing mean survival times for the cancer types. (For this exercise, USE the log transform the data.) For full credit, show the appropriate output.

(a) (2 points) Calculate and list the mean survival time in days for each cancer type using the log function.

(b) (2 points) Check the requirements that the residuals are normally distributed and equal variances.

(c) (6 points) Test the hypothesis

H0: All cancer types have the same mean survival time.

1. State the null and alternative hypotheses,
2. **Ho: All of the population means are equal**
3. **Ha: At least one of the population means is different**
4. Give the ANOVA table
5. Give and Interpret the F statistic
6. 3.871

**The ratio of between group variation vs chance error is 3.871. It appears that we might have a difference for at least one of the groups**

1. Give and interpret the p-value

0.00434

**The probability of calculating the test statistic we got or more extreme assuming the null hypothesis is true is .00434.**

1. State the degrees of freedom for the F statistic,
2. **Degrees of freedom for numerator = 5 Degrees of Freedom for Denominator = 57**
3. Make a conclusion.
4. **The F-statistic is large and the p-value is small, we would reject the null hypothesis. We would then say that at least one of the means is different**

(d) (2 points) Compare your results to problem #3. What are the similarities and/or differences?

**We get the same conclusions stating that we believe that at least one of the means is different. When we transform the data, the F statistic is a little smaller and the p-value is a little larger. However, since all the assumptions are met when we do a transformation, we feel better about using ANOVA with the transformed data.**

(e) (2 points) What are the reasons we want to use replication (more than one unit per treatment group) in a study like this?

**Replication gives us information about the error which then allows us to make inference about the model parameters.**

**DON’T FORGET TO SCORE YOUR HOMEWORK AT THE TOP**