

# QSCI 381 HW 4

Simon-Hans Edasi

5/11/2023

1. (2 points) To what does the word treatment refer to in the ANOVA? Give an example.

**A treatment is a combination of factors being studied. An example would be in a study of interactions between income and gender. Organizing the data into factors of low, middle, and high income; and male, female, and transgender, each combination of income and gender is a treatment.**

2. (4 points) What are the assumptions underlying the ANOVA?

- (a) **The data consists of independent random samples**
- (b) **There should be no significant outliers**
- (c) **The data should be approximately normally distributed**
- (d) **The treatment variances are equal**

3. (1 point) In the ANOVA what are we partitioning?

**ANOVA partitions variances into in-group variance and between-group variance.**

4. (2 points) State the mathematical identity that represents this partitioning in both words and mathematical symbols.

**The variances are partitioned using the total sum of squares identity.**

$$\text{SST} = \text{SSTr} + \text{SSE}$$

5. (6 points) Define each term in the identity above and clearly state what each term measures.

**The total sum of squares identity states that the total sum of squares (SST) is equal to the sum of squares for treatments (SSTr) plus the sum of squares for error (SSE). SSTr measures the between-group variability and SSE measures in-group variance.**

6. (1 point) Describe in words the question the ANOVA is used to answer.

**ANOVA is used to determine if there are statistically significant differences of the means of two or more groups, or if the differences of mean are due to random chance.**

7. (3 points) Clearly state the hypotheses that we are testing in the ANOVA in both words and mathematical symbols.

$$H_o : \mu_1 = \mu_2 = \mu_3 \dots = \mu_n$$

**$H_o$ : The null hypothesis states that the mean of group 1 is equal to the mean of group 2 is equal to the mean of group 3 is equal to the mean of the nth group**

$$H_a : \mu_1 \neq \mu_2 \stackrel{?}{=} \mu_3 \dots \stackrel{?}{=} \mu_n$$

**$H_a$ : The alternative hypothesis states that the means are not all equal.**

8. (2 points) Define the treatment mean square or MS(Tr) in words (what does it measure) and mathematically. Clearly state the probability distribution of this random variable with the proper degrees of freedom.

$$MS(Tr) = \frac{SS(Tr)}{\#Treatments - 1}$$

**The treatment mean square is calculated by dividing the sum of squares by degrees of freedom. This measures variability between group means and represents the average squared difference between group means and overall mean. The MS(Tr) follows the F-distribution with degrees of freedom df1 = # Treatments -1 and df2 = # Treamtents \* (# Observations - 1).**

9. (2 points) Define the error mean square or MSE in words (what does it measure) and mathematically. Clearly state the probability distribution of this random variable with the proper degrees of freedom.

$$MSE = \frac{SSE}{\#Observations - \#Treatments}$$

**The error mean squared (MSE) is a measure of average in-group variability. It is calculated with the sum of squares for error divided by the degrees of freedom. It follows a chi-squared distribution, with DoF = # observations - # treatments.**

10. (2 points) Define the F test statistic in both words and mathematically (include the proper probability distribution for this random variable and degrees of freedom).

$$F = \frac{MS(Tr)}{MSE}$$

**The F-test statistic is used to determine the significance of differences of mean. The F-test statistic follows an F-distribution with df1 = # treatments -1, df2 = # observations - 1.**

11. (1 point) For a given  $\alpha$  -level, state the decision rule (test criterion) for the ANOVA.

**If the calculated F-statistic is smaller than the critical value, then we reject the null hypothesis.**

12. (9 points) Complete the following ANOVA summary table. You may simply use the correct abbreviations instead of the mathematical formulas.

Source of Variation	Degrees of Freedom (df)	Sum of Squares	Mean Square	F value
Treatments	df(Tr)	SS(Tr)	SS(Tr) / df(Tr)	MS(Tr) / MSE
Error	df(Err)	SSE	SSE / df(Err)	*
Total	n-1		*	*

13. (1 point) If we reject  $H_0$  in the ANOVA, we only know that at least one of the treatment means is statistically significantly different from the others. If we wish to identify which of the means are statistically significantly different, why can't we simply perform a series of t-tests on the combinations of all possible pairs of means?

**T-tests only account for comparisons of two groups, and errors can accumulate when performing multiple comparisons. Tukey HSD takes care of this error accumulation.**

14. (28 points) Food processing companies know that peanuts (for making peanut butter) may be contaminated by a fungus, *Aspergillus flavus*, that produces an extremely deadly toxin, aflatoxin. Food inspectors took random samples of the peanuts from three different processing facilities of a food processing company and analyzed the samples for their aflatoxin content (in parts per billion). The following data was recorded:

The inspectors wish to test at the 1% level of significance whether the differences among the three-population means can be attributed to chance.

- (a) (3 points) What three (3) assumptions must be made in order to validly conduct this test?

**The data are normally distributed, free of outliers, and equally varied.**

- (b) (2 points) Clearly state  $H_0$  and  $H_A$  mathematically (please identify all symbols you use).

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : \mu_1 \neq \mu_2 \stackrel{?}{=} \mu_3$$

$\mu_n$  = mean fungus found at  $n$  facility.

- (c) (1 point) Given the context of the situation, why might the inspectors wish to use 1% rather than 5% in performing the analysis? (In answering this question, you may wish to recall the definition of  $\alpha$ .)

**We want to be extra confident because this food can kill people.**

- (d) (8 points) Perform an ANOVA test on these data. Calculate the appropriate sums of squares, degrees of freedom, and mean squares. You may use a calculator or spreadsheet, but not R.
- (e) (9 points) From the analysis, the inspectors calculate  $SST = 336$  and  $SSE = 57$ . Complete the following summary table. Show your work to calculate the degrees of freedom.
- (f) (2 points) Clearly define and give the value of both the test statistic and the critical value. Use the table in the text to determine the critical value and state the appropriate degrees of freedom.
- (g) (2 points) Clearly state the decision rule or test criterion both in terms of the rejection region and the p-value.
- (h) (1 point) State your conclusion in the context of the problem.