Assignment #3: ANOVA

1. An ANOVA was conducted to compare multiple group means. Fill-in the 6 blanks in the following table:

Source of variation	**SS**	**df**	**MS**	**MSR (F)**
Treatments (among groups)	_____	3	158.96	_____
Error (within groups)	_____	32	62.81	(blank)
Total variation	_____	_____	(blank)	η^2 = _____

Based on the above ANOVA summary table:

a. There were _____ groups (treatments) in this study.

b. There were _____ total observations (subjects) in this study.

c. The total variance of all observations in this study was equal to _____

2. An ANOVA was conducted to compare multiple group means. Fill-in the 6 blanks in the following table:

Source of variation	**SS**	**df**	**MS**	**MSR (F)**
Treatments (among groups)	_____	7	_____	5.01
Error (within groups)	_____	_____	3.62	(blank)
Total variation	_____	29	(blank)	η^2 = _____

Based on the above ANOVA summary table:

a. There were _____ groups (treatments) in this study.

b. There were _____ total observations (subjects) in this study.

c. The total variance of all observations in this study was equal to _____

Scenario: A sample of 20 different types of cereals was taken from each of three grocery store shelves (the lowest, middle, and highest shelves). A summary of the sugar content (grams per serving) and dietary fiber (grams per serving) of the cereals is given below.

	Sugar content (g per serving)		
	n	$\overline{X}$	**s**
Lowest shelf	n = 20	4.80	2.138
Middle shelf	n = 20	9.85	1.985
Highest shelf	n = 20	6.10	1.865
All shelves combined	N = 60	M = 6.9167	s = 2.9194

3. In order to conduct an ANOVA, we need to make some assumptions: independence, normality, & equal variances. Determine whether the equal variance assumption is satisfied for this study by conducting an Fmax test.

For the sugar study: Fmax = _____. The equal variance assumption **IS** **IS NOT** satisfied.

4. We want to determine if the (population) mean sugar levels differ among the 3 shelves. What are the hypotheses?

Null hypothesis: _____ Alternate hypothesis: _____

5. For the sugar data, fill-in the following ANOVA summary table by calculating SS, df, MS, MSR, and eta-squared. Just this once, try to calculate all of these things by hand (using a calculator; not a statistical program or applet).

Source of variation	**SS**	**df**	**MS**	**MSR (F)**
Shelves	_____	_____	_____	_____
Error	_____	_____	_____	(blank)
Total	_____	_____	MS_{total}	η^2 = _____

Once you have the table completed, go ahead and check your work with http://danielsoper.com/statcalc3/calc.aspx?id=43

6. Calculate MS_{total} for the sugar data (use the ANOVA summary table you just created). The MS_{total} was actually given to you in the scenario. What does MS_{total} represent?

   MS_{total} represents: _____

7. If the null hypothesis were true, how unlikely were we to get the sugar data that we observed in this study? Use an F-distribution table or calculator to estimate the p-value for your MSR.

   Calculator: http://lock5stat.com/statkey/theoretical_distribution/theoretical_distribution.html#F
   F-table: http://bradthiessen.com/html5/stats/m301/ftable.pdf

   Identify the degrees-of-freedom for your F-statistic (MSR): Numerator df = _____ Denominator df = _____

   Approximate p-value: $P(F \geq MSR)$ =_____

8. From this p-value write out any conclusion(s) you can make for this sugar study:

   Conclusions: _____

9. Look at the eta-squared you calculated in the ANOVA summary table. Interpret this eta-squared with regards to this study.

   Interpretation: _____

Scenario: Suppose we're interested in testing the critical thinking skills of seniors at St. Ambrose. We randomly sample seniors majoring in each of the following 6 groups of programs:

- STEM majors (e.g., math, chemistry, engineering, computer science)
- Health science majors (e.g., biology, nursing, exercise science)
- Social science majors (e.g., psychology, sociology, criminal justice, political science)
- Humanities majors (e.g., English, philosophy, theology, history)
- Creative arts majors
- Education majors

We then administer a test of critical thinking to these students.

10. We want to compare the means of 6 different groups. To do this, we could decide to conduct a series of t-tests to compare all possible pairs of group means. Answer the following:

    a. How many t-tests would we need to conduct to compare all possible pairs of our 6 group means? _____

    b. Suppose we set $\alpha = 0.05$ for each of those t-tests. What would be the probability of making at least one α-error across all our t-tests?

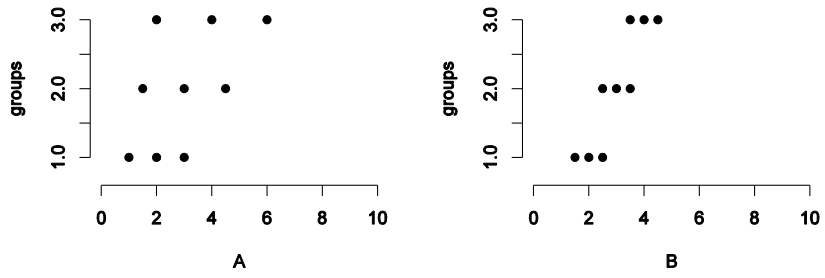
    Answer: _____

    c. Suppose you conduct an ANOVA to compare the 6 group means. What would happen to your F-statistic (MSR) and the p-value under each of these scenarios:

    i. If we sample a larger number of students in each group, the F-statistic will **INCREASE** **DECREASE**

    ii. If we sample a larger number of students in each group, the p-value will **INCREASE** **DECREASE**

    iii. If the population (group) means remain the same but the group standard deviations all increase, the F-statistic will: **INCREASE** **DECREASE**



11. Datasets A and B are displayed above. Use that pair of plots to answer the following:

    a. If you conducted an ANOVA for each dataset, which one would have the larger MSA (between groups)?

    Circle your answer: **DATASET A** **DATASET B** **They are equal** **Unable to determine**

    b. If you conducted an ANOVA for each, which dataset would have the larger MSE (within groups)?

    Circle your answer: **DATASET A** **DATASET B** **They are equal** **Unable to determine**

    c. Based on the MSR (F-statistic) you would calculate for each, which would yield the smaller p-value?

    Circle your answer: **DATASET A** **DATASET B** **They are equal** **Unable to determine**

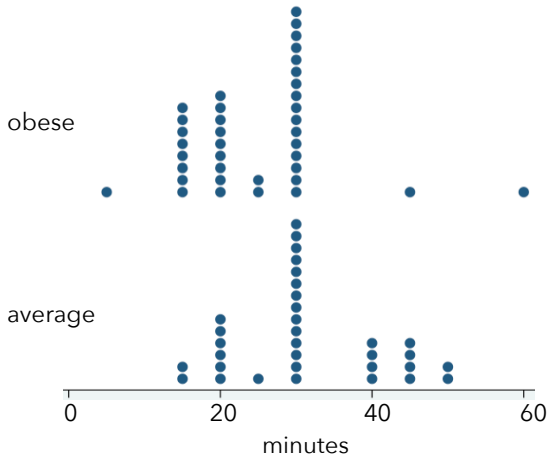
Scenario: Does the amount of time doctors spend with patients depend on whether the patient is obese?

A 2001 study examined physicians' behavioral intentions as well as their expressed attitudes towards average-weight and obese patients. 71 primary care physicians in Houston participated in this study. The doctors were sent a packet containing a medical chart similar to the one they view upon seeing a patient. This chart portrayed a patient who was displaying symptoms of a migraine headache but was otherwise healthy. The weight of the patient was manipulated so that:

- 33 doctors received a chart from a patient of average weight (body mass index = 23)
- 38 doctors received a chart from an obese patient (body mass index = 36)

The doctors were instructed to examine the charts and then asked, among other questions, how much time they believed they would spend with the patient. Here's a summary of that data:

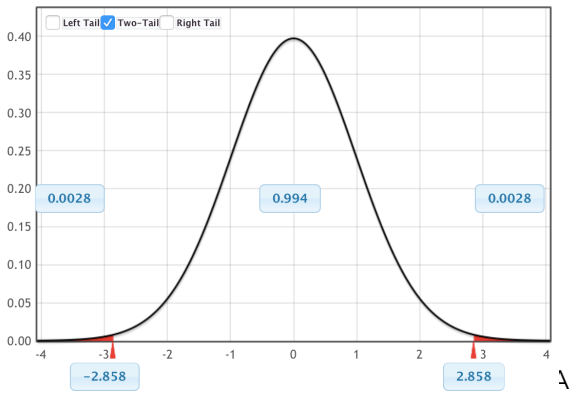
weight	N	mean	p50	sd
average	33	31.36364	30	9.864134
obese	38	24.73684	25	9.652571
Total	71	27.8169	30	10.23762



12. If we were to conduct an independent samples t-test (assuming the population variances are equal), we would calculate an observed t-statistic of:

$$t_{33+38-2} = \frac{31.36364 - 24.73684}{\sqrt{\frac{1}{33}+\frac{1}{38}}\sqrt{\frac{(33-1)(9.864134)^2+(38-1)(9.652571)^2}{33+38-2}}} = \frac{6.6268}{(0.23795)(9.75125)} = 2.856$$

If we were running a two-tailed test (i.e., our alternative hypothesis was $\mu_{average} = \mu_{obese}$), we would estimate the p-value as p= 0.0028 x 2 = 0.0057 (as displayed in the t-distribution to the right)



If we conducted an ANOVA on this same data (to compare the two group means), what would we calculate for our F-statistic (our MSR)? What would our p-value be?

If you don't know, you could always type the sample sizes, means, calculator: http://danielsoper.com/statcalc3/calc.aspx?id=43

The F-statistic (MSR) would equal: _____ (look at how that compares to the t-statistic calculated above).

The p-value would be: _____.