*YOUR TA’S NAME*:

*Lecture Worksheet*

*Tuesday 10/27/2020*

**MAIN POINTS OF LECTURE**

1. **AN**alysis **O**f **VA**riance (**ANOVA**) techniques are used to compare the mean of some continuous variable *Y* across *J* populations using data from independent samples from each of the *J* populations. This amounts to comparing the mean of some continuous variable *Y* across the *J* categories of some discrete variable *X*
2. ANOVA amounts to a test of the hypothesis that all of the *J* population means are equal:
	* 1. H0: **1=**2=**3=...=**j
		2. H1: Not all of the means are equal
	1. If we reject the null, then we reject the claim that continuous variable *Y* and discrete variable *X* are not associated in the population
3. The test statistic that we use to perform the ANOVA hypothesis test is called the *F* statistic. Conceptually, the more the *J* sample means differ, the larger the *F* statistic. If *F* is larger than we would expect by chance … if the observed value of the *F* test statistic exceeds a critical value determined in advance … then we reject H0
4. The *F*-statistic is based on a comparison of between-group variability to within-group variability

$$F\_{J-1,n-J}=\frac{\sum\_{j=1}^{J}n\_{j}(\overline{Y}\_{j}-\overline{Y})^{2}/(J-1)}{\sum\_{j=1}^{J}\sum\_{i=1}^{n\_{j}}(Y\_{ij}-\overline{Y}\_{j})^{2}/(n-J)}=\frac{SS\_{BETWEEN}/J-1}{SS\_{WITHIN}/n-J}=\frac{MS\_{BETWEEN}}{MS\_{WITHIN}}$$

1. MSBETWEEN is an estimate of the amount of variance in *Y* attributable to the category of *X* to which cases belong
2. MSWITHIN is an estimate of the amount of variance in *Y* attributable to error (or random variability)
3. MSBETWEEN has *j* — 1 degrees of freedom and MSWITHIN has N — *j* degrees of freedom … consequently, F thus has two degrees of freedom (*df*): The “numerator” *df*NUM (1=*j*—1) and the “denominator” *df*DENOM (2=N—*j*)

**QUESTIONS**

1. [From recorded lecture] With *a*=0.05, what is the critical value of *F* when…

dfNUM=3, and dfDENOM=120 ? **2.70**

dfNUM=10, and dfDENOM=20 ? **2.35**

dfNUM=20, and dfDENOM=30 ? **1.93**

1. [From recorded lecture] With *a*=0.01, what is the critical value of *F* when…

dfNUM=1, and dfDENOM=1,000 ? **6.63**

dfNUM=4, and dfDENOM=20 ? **4.43**

1. [From recorded lecture] There are three delivery companies: A, B, & C. I had all three mail me 5 packages. Below are the number of days it took for me to get the packages

 A 2 2 3 4 6 **Hint/Help**:

 B 1 2 2 5 5 SSBetween = 0.9333

 C 2 2 3 3 4 SSWithin = 28

Test the hypothesis that the mean number of days that each company takes to deliver packages is equal; use =0.05 (For this example, relax/ignore the basic assumptions that must be met in order to perform ANOVA)

H0: **1=**2=**3

H1: Not all of the means are equal

J=3 and N=15, so critical value of F2,12 with a=0.05 is 3.89

$F\_{2,12}=\frac{SS\_{BETWEEN}/J-1}{SS\_{WITHIN}/n-J}=\frac{0.9333/2}{28/12}$ = 0.2

Since our F statistic (0.2) does not exceed the critical value of F (3.89), we fail to reject H0.

1. [From recorded lecture] There are three delivery companies: A, B, & C. I had all three mail me 3 packages. Below are the number of days it took for me to get the packages

 A 1 1 2 **Hint/Help**:

 B 2 2 3 SSBetween = 6

 C 3 3 4 SSWithin = 2

Test the hypothesis that the mean number of days that each company takes to deliver packages is equal; use =0.05 (For this example, relax/ignore the basic assumptions that must be met in order to perform ANOVA)

H0: **1=**2=**3

H1: Not all of the means are equal

J=3 and N=9, so critical value of F2,9 with a=0.05 is 4.26

$F\_{2,9}=\frac{SS\_{BETWEEN}/J-1}{SS\_{WITHIN}/n-J}=\frac{6/2}{2/6}$ = 9

Since our F statistic (9) exceeds the critical value of F (4.26), we reject H0.