TA’S NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem Set #6**

1. Report the critical values of *F* for the following:
	1. =0.05, 2 groups, 20 total subjects ANSWER: 4.41
	2. =0.05, 4 groups, 30 total subjects ANSWER: 2.98
	3. =0.01, 4 groups, 30 total subjects ANSWER: 4.64
	4. =0.01, 10 groups, 100 total subjects ANSWER: 2.65
	5. =0.05, 9 groups, 1,000 total subjects ANSWER: 1.98
2. How many times have Chevy, Ford, and Dodge truck drivers had their trucks break down? From within each of the *j*=3 groups, I randomly sampled 3 drivers. The resulting responses were:

Among CHEVY Drivers: 1, 1, 3

Among FORD Drivers: 0, 1, 4

Among DODGE Drivers: 1, 2, 5

SSBetween = 2

SSWithin = 20

Conduct—by hand—an analysis of variance to determine whether there is a statistically significance association between Y=”how many times drivers’ trucks have broken down” and X=”the make of the truck.” (*For this example, relax the assumptions about the normality and homoskedasticity assumptions*). Use =0.05. Be sure to (a) State the hypotheses; (b) State the critical value; (c) Calculate and report the appropriate test statistic; and (4) State your conclusions based on a comparison of the test statistic to the critical value. *Show your work.*

H0: CHEVY=FORD=DODGE; H1: Not all three means are equal

With =0.05, N=9, and j=3, dfNUM=3-1=2 and dfDENOM=N-j=6, so the critical value of F equals 5.14

$$F\_{J-1,n-J}=\frac{SS\_{BETWEEN}/J-1}{SS\_{WITHIN}/n-J}=\frac{2/2}{20/6}=0.3$$

Because the value of the test statistic (0.3) does not exceed the critical value (5.14), we fail to reject H0. That is, we conclude that the mean number of breakdowns does not vary by truck type.

1. Do college students’ grade point averages vary by how many hours per week they work at paid jobs? To find out, I randomly selected four students who do not hold paid jobs during the academic year, four students who work at part-time jobs during the school year, and four students who work at full-time jobs during the school year. I collected information about all 12 students’ grade point averages. Here is what I observed:

*No Paid Jobs*:

Student #1: 3.6 Student #2: 3.1 Student #3: 3.3 Student #4: 3.2

*Part-Time Jobs*:

Student #1: 3.3 Student #2: 3.4 Student #3: 3.0 Student #4: 3.1

*Full-Time Jobs*:

Student #1: 3.5 Student #2: 2.8 Student #3: 3.1 Student #4: 3.0

SSBetween = 0.08

SSWithin = 0.50

Test the hypothesis that the population mean GPA is the same across groups of students defined by whether they work full-time, part-time, or not at all at paid jobs during the school year. (*For this example, relax the assumptions about the normality and homoskedasticity assumptions*). Use =0.05. Be sure to (a) State the hypotheses; (b) State the critical value; (c) Calculate and report the appropriate test statistic; and (4) State your conclusions based on a comparison of the test statistic to the critical value. *Show your work.*

H0: **1=**2=**3 (mean GPA is the same across groups of students)

Ha: Not all of the three population means are the same

With =0.05, N=12, and j=3, dfNUM=3-1=2 and dfDENOM=N-j=9, so the critical value of F equals 4.26

$$F\_{J-1,n-J}=\frac{SS\_{BETWEEN}/J-1}{SS\_{WITHIN}/n-J}=\frac{0.08/2}{0.50/9}=0.72$$

Because F does not exceed 4.26, we fail to reject the null hypothesis that all the means are equal. Our results do not provide evidence that population mean GPAs differ across groups of students defined by whether they work full-time, part-time, or not at all at paid jobs during the school year.

1. Report the critical values of 2 for the following tables with *i* rows and *j* columns:
	1. =0.05, *i*=2 and *j*=4 ANSWER: 7.815
	2. =0.05, *i*=3 and *j*=6 ANSWER: 18.307
	3. =0.01, *i*=3 and *j*=2 ANSWER: 9.210
	4. =0.01, *i*=5 and *j*=9 ANSWER: 50.892
	5. =0.05, *i*=10 and *j*=3 ANSWER: 28.869
2. A group of researchers is interested in understanding how television viewing among children is related to their grades in school. They interviewed 200 randomly selected children and asked them how many hours of television they typically watch per day. They then found out from the children’s parents what their grades are in school. Here are their results:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **GRADES** |  |
|  |  | Mostly A’s & B’s | Mostly C’s | Mostly D’s or F’s | *Total* |
| **HOURS OF TELE-VISION WATCHED PER DAY** | None | 60 | 13 | 7 | 80 |
| *48* | *24* | *8* |  |
| <2 Hours | 45 | 37 | 8 | 90 |
| *54* | *27* | *9* |  |
| >2 Hours | 15 | 10 | 5 | 30 |
| *18* | *9* | *3* |  |
|  |  |  |  |  |  |
|  | *Total* | 120 | 60 | 20 | 200 |

Is there a statistically significant association between students’ grades and the number of hours of television per day that they watch? Use  significance level of 0.05.

* 1. State the null and alternative hypotheses

ANSWER:

H0: There is no association between grades and number of hours of TV watched per day

Ha: There is an association between these two variables

* 1. Do these sample data conform to the basic assumptions that must hold in order to make valid inferences using this procedure?

ANSWER: Yes, the children were selected at random and there are at least 5 observations in each cell.

* 1. Determine the “critical value” … that is, how large the test statistic must be in order to reject the null hypothesis at the given  level

ANSWER: w/ =0.05 and *df*=(3-1)(3-1)=4, the test statistic 2 must exceed 9.488 in order to reject H0

* 1. Compute the appropriate test statistic. *Show your work.*

ANSWER:

The first step is to compute the expected values for each cell (which I have inserted into the table), where .

Next, we compute the test statistic:

=3.00+5.04+0.13+1.50+3.70+0.11+0.50+0.11+1.33=**15.43**.

* 1. Compare the test statistic to the critical value. What do you conclude? Do you reject the null hypothesis? In plain terms, explain what your statistical conclusion says about whether there is a significant association between students’ grades and the number of hours of television per day that they watch.

ANSWER: We reject the null hypothesis because 2 exceeds 9.488. This means that we conclude that there is a statistically significant association between grades and number of hours of television watched per day.

1. Based on the sample data described in the crosstable below, is there a statistically significant association between X and Y in the population from which these data were randomly selected?

Y=1 Y=2

X=1 10 5

X=2 7 5

X=3 5 10

Conduct—by hand—a 2 analyses to determine whether there is a statistically significance association between X and Y. Use =0.05. Be sure to (a) State the hypotheses; (b) State the critical value; and (c) State your conclusions based on a comparison of the test statistic to the critical value. *Show your work.*

ANSWER:

1. H0: X and Y are statistically independent; H1: X and Y are not statistically independent
2. With a 3x2 table and =0.05, the critical value is 5.991. Thus we will reject H0 if 2 exceeds 5.991
3. To compute 2 we must first compute the expected values for each cell:

Y=1 Y=2 TOTAL

Observed: X=1 10 5 15

Expected (15x22)/42=7.86 (15x20)/42=7.14

Observed: X=2 7 5 12

Expected (12x22)/42=6.29 (12x20)/42=5.71

Observed: X=3 5 10 15

Expected (15x22)/42=7.86 (15x20)/42=7.14

Observed Total 22 20 42

2 then equals: (7.86-10)2/7.86+(7.14-5)2/7.14+(6.29-7)2/6.29+(5.71-5)2/5.71+ (7.86-5)2/7.86+(7.14-10)2/7.14=0.605+0.641+0.080+0.088+1.041+1.146=3.601

Because the value of the test statistic (3.601) does not exceed the critical value (5.991), we fail to reject H0. That is, we conclude that X and Y are statistically independent.

1. Below is a table relating the dichotomous variable “Happy or Sad” to the dichotomous variable “In Love or Not.” Describe the strength and direction of the association between these two variables using **(a)** Relative Risk and **(b)** the Odds Ratio (each of which you should compute by hand). For the latter measures, consider “In Love or Not” to be the dependent variable. Be sure to interpret each value. **SHOW YOUR WORK**

**Happy or Sad?**

*Sad Happy*

 **In Love or** *In Love* 11 18

 **Not?** *Not in Love* 21 14

(a) Relative Risk=1.64. That is, happy people are at 64% greater risk of being in love.

 (b) Odds Ratio = 2.45. That is, the odds of being in love are 145% greater among happy people than among sad people.