Edhesive AP Statistics **Unit 1 – Solutions**

**Multiple Choice:** Choose the best answer choice for the following problems.

*The following situation applies to questions 1-5*

A number of SRSs were conducted in order to determine if there were a difference in the presidential approval rating between households that have an income above $250,000/yr versus those below this income level. Of the 1000 households above $250k/yr that were surveyed, 380 approved, while of the 1200 households below $250k/yr, 400 approved.

1. Let be the proportion of households with income above $250k/yr that approve and the proportion below $250k/yr that approve. Given what is stated above, which is the correct null/alternative hypotheses to test?
   1. versus
   2. versus
   3. versus
   4. versus
   5. versus

The null hypothesis in this case would be that there is no difference in approval rating between these two groups (). The alternative is that there is a difference ().

1. The researchers report that the test is significant at the 5% level, but not at the 1% level. Which statement is the most accurate summary of the results?
   1. The P-value is between 0.01 and 0.05, thus we can reject the null hypothesis at the 95% level but not the 99% level and say with a confidence level of 95% that there is a difference in approval rating between the two groups.
   2. The P-value is less than 0.01, thus we can reject the null hypothesis at the 99% level but not the 95% level and say with a confidence level of 99% that there is a difference in approval rating between the two groups.
   3. The P-value is greater than 0.05, thus we cannot reject the null hypothesis at the 95% level but can at the 99% level and say with a confidence level of 99% that there is a difference in approval rating between the two groups.
   4. The P-value is between 0.01 and 0.05, thus we cannot reject the null hypothesis at the 95% level but can at the 99% level and say with a confidence level of 99% that there is a difference in approval rating between the two groups.
   5. The P-value is between 0.01 and 0.05, thus we can reject the alternative hypothesis at the 95% level but not the 99% level and say with a confidence level of 95% that there is no difference in approval rating between the two groups.

Meeting (and failing to meet) significance at the 95% (99%) level means the P-value is <0.05 (>0.01). Thus we know that 0.01<P-value<0.05. This means that at the 95% confidence level we can reject the null hypothesis and say that there is a difference between the two groups. At 99% we cannot reject the null hypothesis and we cannot say if there is a difference.

1. Achieving a confidence level of 95% but not 99% in this context means that the test resulted in a z-value of-

For a 2-tailed test, P-value=1-2\*(z-table value). In this case then we look for z-table values >0.475 and <0.495. This occurs for 1.96<z<2.58.

1. In order to publish the results of the study the researchers must include the margin of error. Which of the following represents the correct formula to calculate the margin of error for a 95% confidence interval?

For a two population margin of error the formula is where is the z-value corresponding to the desired confidence level (in this case for a 95% confidence interval) and represents the proportion of groups 1,2 with sample sizes of .

1. To see if your neighbourhood agrees with the results of this study, you survey some of your neighbors. Of the 20 household above $250k/yr, 6 approved, whereas of the 15 households below $250k/yr only 3 approved. Which of the following best describes your test?
   1. , my test does not suggest a difference at a 95% confidence level.
   2. , my test suggests a difference at a higher level of confidence than the study.
   3. , my test suggests a difference at the same level of confidence as the study.
   4. Inconclusive- the counts for households below $250k/yr is too low.
   5. Inconclusive- the random condition is not met.

In this test we require and for both populations. The >250k/yr population meets this requirement (np=20\*6/20=6 ; n(1-p)=20(1-6/20)=14) whereas the <250k/yr sample does not (np=15(3/15)=**3** ; 15(1-3/15)=12).

**Free Response – Solutions**

1. You happen to be a fan of the Lucky Charms cereal, particularly the 30% of the pieces which are the marshmallow rainbows, hearts, moons, etc. A competing brand, “Fortunate Totems” contains only 27% marshmallow pieces. Suppose you take random samples of 100 pieces from each cereal brand and calculate (the proportion of marshmallows in Lucky Charms) and (the proportion of marshmallows in Fortunate Totems).
   1. Describe the sampling distribution of .

The sample mean is

The standard deviation is

Since and for the Lucky Charms (Fortunate Totems) sample, all of which are at least 10, the distribution is approximately Normal.

* 1. What is the probability that your sample of Fortunate Totems contains more marshmallow pieces than your sample of Lucky Charms?