Edhesive AP Statistics **Unit 10, Lessons 1 & 2– Solutions**

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

*Questions 1-5 apply to the following situation*

A local author wants to create a character for a new children’s book and wants this character to have some sort of super power. To make the book more appealing to the kids in the neighborhood she starts to randomly survey kids to see which super power they would prefer whenever she goes to the park. Below are her data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Super Power** | Flight | X-ray Vision | Invisibility | Regeneration | Telekinesis | **Total** |
| **Count** | 51 | 32 | 29 | 28 | 45 | 185 |

1. In order to determine the statistical “winner” among the super powers the author needs help formulating the right test. Of the following options, which combination will create an appropriate null hypothesis for this test?
2. Where represents the number of kids in the neighborhood that would choose Flight as their preferred super power, and so on.
3. Where represents the proportion of kids in the sample that chose Flight as their preferred super power, and so on.
   1. I and III
   2. I and IV
   3. II and III
   4. II and IV
   5. None of these pairs create an appropriate null hypothesis

The null hypothesis in this case is that all super powers are preferred equally, thus F=X=I=R=P is appropriate. Each of these letters should then represent the proportion of the *sample* that favor the specific super power. Therefore I and IV create an appropriate null hypothesis.

1. Based on the data she collected, how should she compute the statistic?

We use the formula where O=observed and E=expected and the sum is taken over all parameters.

1. Assuming 4 degrees of freedom and a chi-squared value of 11.62, which of the following is true?

* 1. P-value<0.01
  2. 0.01<P-value<0.05
  3. 0.05<P-value<0.1
  4. 0.1<P-value<0.25
  5. Not enough information

Using chi-squared distribution table, the value of 11.62 occurs between 0.01 and 0.05 when there are 4 degrees of freedom.

1. If the author conducted her test at a level of , what is the appropriate conclusion to draw from the test?

* 1. Because the P-value > 1% she cannot reject the null hypothesis. There is not sufficient evidence to claim that there are differences in the kids’ super power preferences.
  2. Because the P-value > 1% she can reject the null hypothesis and claim that there are statistically significant differences in the kids’ super power preferences.
  3. Because the P-value < 5% she can reject the null hypothesis and claim that there are statistically significant differences in the kids’ super power preferences.
  4. Because the P-value < 5% she cannot reject the null hypothesis. There is not sufficient evidence to claim that there are differences in the kids’ super power preferences.
  5. Because the P-value > 1% she cannot reject the null hypothesis. There is no difference in the kids’ super power preferences.

Given that the desired level of confidence was , but the P-value was > 0.01, there is insufficient evidence to reject the null hypothesis. Therefore we cannot claim statistically significant differences in the kids’ super power preferences.

1. The author conducts a similar test of versus to determine if the kids prefer the ability to fly over the other 4 super power options and finds a P-value of 0.10. Assuming she writes her new book about a girl who can fly, what is the likelihood that in reality the kids in the neighborhood do not prefer flight as their chosen super power?

* 1. 10%
  2. 90%
  3. 0.1%
  4. 1%
  5. Not enough information

This represents a Type I error: the incorrect rejection of a true null hypothesis. In this case a P-value of 0.10 means there is a 10% chance that a study would produce as or more extreme results under the assumption of the null hypothesis.

**Free Response – Solutions**

1. Bill had been training for a cycling race for a few months when he has a crash. Jill, Bill’s cycling coach, is concerned that he is now biking more slowly. To measure this, she compares the number of minutes Bill spends in each of a few workout zones as determined by his heart-rate. Before his accident, Bill spend 10% of his time in the Low zone, 15% in Medium, 45% in High, and 30% in the Extreme zone. After the accident she records his performance over a few weeks of training recording the number of minutes spent in each zone. Below are her results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Low** | **Medium** | **High** | **Extreme** | **Total** |
| **Minutes in Zone** | 35 | 55 | 98 | 52 | 240 |

* 1. If we wish to perform a significance test to determine if the distribution of time spent in each zone has changed, what are the appropriate null and alternative hypotheses to test? Are the conditions met for this test?

The distribution of time spent in each zone did not change after the accident

The distribution of time spent in each zone changed after the accident

Conditions:

Random- Assuming the minutes recorded represent a random sample of Bill’s cycling performance.

10%- Assuming Bill has cycled for at least 2400 minutes

Large counts- All the counts are >5

* 1. Assuming the null hypothesis, compute the expected time in each zone.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Low** | **Medium** | **High** | **Extreme** |
| **Expected Minutes in Zone** | 0.1\*240=24 | 0.15\*250=36 | 0.45\*240=108 | 72 |

* 1. Compute the test statistic and P-value for this test and draw an appropriate conclusion using a confidence level of .

df=3; ; so P-value <0.01. Therefore we can reject the null hypothesis at the 99% confidence level and say that Bill’s distribution of time spent in each zone has changed since his accident.

* 1. Based on your conclusion in part (c), which type of error (I or II) is it possible you have made?

It is possible we have made a Type I error, incorrectly rejecting a true null hypothesis.

* 1. Calculate the components of the chi-squared statistic and comment on which zones were most affected by Bill’s accident.

The components of the chi-squared statistic are

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Low** | **Medium** | **High** | **Extreme** |
| **Chi-square component** |  |  |  |  |

The observed number of minutes in the Medium zone was much higher than expected, primarily due to a much lower number of observed minutes in the Extreme zone.