Edhesive AP Statistics **Unit 10, Lesson 3– Solutions**

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

*Questions 1-5 apply to the following situation*

In order to assess the demographic trend surrounding the debate around marijuana legalization, a study interviewed a random sample of 1,808 people below the age of 35 and 1,763 people over the age of 35 from around the country and asked them to choose which opinion most closely matched their own. Here are the results:

|  |  |  |  |
| --- | --- | --- | --- |
| **Opinion** | **Age < 35** | **Age > 35** | **Totals** |
| I favor total legalization | 534 | 97 | 631 |
| I favor legalization but with restrictions and limitations | 422 | 302 | 724 |
| I favor medical legalization but not recreational use | 415 | 365 | 780 |
| I favor almost no legalization, with few exceptions for extreme cases | 125 | 482 | 607 |
| I favor no legalization | 312 | 517 | 829 |
| **Totals** | 1808 | 1763 | 3571 |

1. When designing a chi-squared test of homogeneity for this data, what is the appropriate null hypothesis to test against?
   1. The same proportion of people in both age groups favor total legalization
   2. There is no association between age and opinions about legalization in the population
   3. There is no difference between the distributions of the two age groups’ opinions about legalization in the population
   4. There is no association between age and opinions about legalization in the sample
   5. There is no difference between the distributions of the two age groups’ opinions about legalization in the sample

A chi-squared test of homogeneity is designed to attempt rejection of the null hypothesis that the data represent the same distribution. In this case, that those surveyed below 35 years old have a distribution of opinions the same as those above 35 years old.

1. What is the expected count of people over the age of 35 that support no legalization?
   1. 357.1
   2. 352.6
   3. 409.3
   4. 414.5
   5. 517.0

When testing for homogeneity, we compute a pooled ‘expected’ proportion based on the total number of people with each opinion divided by the total number surveyed. In the case of the opinion ‘no legalization’ this gives . This, times the total number of people over 35 who were interviewed, gives us an expected value of

1. If the test is to be conducted at a level, what value of will mean there is sufficient evidence to reject the null hypothesis?

Using df=(#rows-1)(#columns-1)=4 we see a value of 9.49 results in a P-value of 0.05.

1. This test results in a value of and a P-value near 0.0. Assuming a significance level of 0.01, which of the following are true?

* 1. Type I error possible and extremely likely
  2. Type II error possible and extremely likely
  3. Type I error possible but extremely unlikely
  4. Type II error possible but extremely unlikely
  5. Both types of errors are possible

A P-value near 0 means we have reached statistical significance at the 0.01 level and can reject the null hypothesis. This means committing a Type I error (incorrect rejection of a true null hypothesis) is possible. However, since the P-value is near zero, it is extremely unlikely that we have done so.

1. Which of the following would *NOT* display information about the differences of opinions between age groups?
2. A bar chart showing the marginal distribution of opinion about marijuana legalization
3. A bar chart showing the marginal distribution of age group
4. A bar chart showing the conditional distribution of age group for each opinion
5. A bar chart showing the conditional distribution of opinion for each age group
   1. All of these display information about differences in opinion by between age groups
   2. IV only
   3. III and IV
   4. I only
   5. I and II

A marginal distribution effectively shows row/column totals. Option I would show how many people from the entire population held each opinion without providing any age group information, while Option II would show how many people from each age group were sampled without showing any information about their distribution of opinions.

**Free Response – Solutions**

1. Stacy works for the state of California and manages all of the food orders for the state’s zoos. In order to try to get a discount from the manufacturer, she wants to use one brand for all of the zoos. She conducts a study in which a SRS of a few types of animals are given different brands of food. The table below shows the number of each type of animal that was found to prefer each type of food.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Zoo Food** | **Safari Nibs** | **Healthy Monkey** | Totals |
| **Lions** | 8 | 10 | 7 | 25 |
| **Hippos** | 23 | 21 | 17 | 61 |
| **Giraffes** | 12 | 15 | 9 | 36 |
| Totals | 43 | 46 | 33 | 122 |

* 1. Fill out the table below with the appropriate conditional distributions based on the data collected above for comparing the food preference of the three different animals.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Zoo Food** | **Safari Nibs** | **Healthy Monkey** |
| **Lions** | 0.32 | 0.40 | 0.28 |
| **Hippos** | 0.38 | 0.34 | 0.28 |
| **Giraffes** | 0.33 | 0.42 | 0.25 |

We compute a conditional distribution using the total for each animal, so for Lions, the conditional proportions are 8/25, 10/25, and 7/25 for each food brand.

* 1. In the space below, present the data from part (a) graphically and comment on the relationship between the type of animal and their food brand preference.

Here are two representations of the data presented in part (a) showing the conditional distribution of brand preference for each animal. It is apparent that for both the Lions and Giraffes, Safari Nibs are the preferred brand while the Hippos prefer Zoo Food. Healthy Monkey ranked lowest for all three groups of animals.

* 1. Conduct the appropriate statistical test to determine if the food brand preference varies between the three types of animals.

We are testing The distribution of food preference is the same for each type of animal, versus : The distribution of food preference is not the same for each animal type. We’ll use a chi-square test of homogeneity at a level of . First we check conditions: *Random*- She collected her data from a SRS of each population. *10%*: It is assumed the sample of each animal represents less than 10% of the total population of that animal in the CA zoo system. *Large Counts*- We compute expected counts as in problem 2 () to get:

|  |  |  |  |
| --- | --- | --- | --- |
| Expected Counts | **Zoo Food** | **Safari Nibs** | **Healthy Monkey** |
| **Lions** | 8.81 | 9.43 | 6.76 |
| **Hippos** | 21.50 | 23.00 | 16.50 |
| **Giraffes** | 12.69 | 13.57 | 9.74 |

Since all expected counts are >5 we satisfy this requirement.

We compute . With this gives . We cannot reject the null hypothesis. There is convincing evidence that the distribution of food preference is not different for the different types of animals.

* 1. If you chose a chi-square test of homogeneity for part (c), explain how the data could have been collected to make a chi-square test for independence. If you chose a chi-square test of independence for part (c), explain how the data could have been collected to make a chi-square test for homogeneity.

We ran a chi-square test of homogeneity to test if the food preference distribution was different for different types of animals. If we wished instead to run a chi-square test of independence a SRS of all zoo animals could have been taken. From this sample we could have then treated the type of animal and food preferred as two categorical variables making a test of independence appropriate.