Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date\_\_\_\_\_\_\_\_\_\_\_\_\_

AP Statistics

Chapter 11 Take Home

1. The HR supervisor at a large company (more than 20,000 employees) believes that people are more likely to call in sick on Monday or Friday, so they can take a long weekend. She took a random sample of 850 reports of used sick days from the past few years and determined the day of the week for each report. Here are the results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Day | Monday | Tuesday | Wednesday | Thursday | Friday |
| # of sick days | 192 | 151 | 148 | 152 | 207 |

(a) Do the data provide convincing evidence that absences for sick days are not evenly distribution through the week? Justify your answer with appropriate statistical evidence.

(b) Based on your answer to (a), which error is it possible that you have made, Type I or Type II? Describe that error in the context of the problem.

(c) Use components of the chi-square statistic to perform a follow-up analysis on whether the data supports the claim that employees may be using sick days to take long weekends.

2. Shoprite sells four different sizes of a popular brand of Cheerios. For the past few years the proportion of boxes they sell of each size has been quite stable: 10% Small, 15% Medium, 60% Large, and 15% Jumbo. They decide to change the pricing of the four sizes and want to see if this changes the proportion of boxes they sell of each size. To test this, a few weeks after changing the prices they take a simple random sample of 120 transactions involving Cheerios and count how many boxes of each size were sold. Here are the results.

|  |  |  |  |
| --- | --- | --- | --- |
| Observed # of boxes sold for each size | | | |
| Small | Medium | Large | Jumbo |
| 11 | 17 | 63 | 29 |

(a) We wish to carry out a test of significance to see if the distribution of sizes of cereal boxes sold has changed. State the null and alternative hypotheses for this test.

(b) Find the expected counts for each size box under the assumption that the null hypothesis is true.

|  |  |  |  |
| --- | --- | --- | --- |
| Expected # of boxes sold for each size | | | |
| Small | Medium | Large | Jumbo |
|  |  |  |  |

(c) Discuss whether the conditions for this test have been met.

(d) Find the value of the test statistic and the *P*-value of the test, and make the appropriate conclusion. Use  = 0.05.

(e) Based on your answer to (d), which error is it possible that you have made, Type I or Type II? Describe that error in the context of the problem.

(f) Use the components of the chi-square statistic to perform a follow-up analysis on the impact of the new prices on the sales of different sizes of cereal boxes.

3. The managers of a radio station catering to young people would like to determine what genres of music they should play in order to maximize the size of their listening audience. They take a simple random sample of 50 students at each of three schools: a local middle school, a high school, and a college. The students are asked to choose which of three different types of music they most enjoy hearing on the radio. Here are the results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Music Genre | | |  |
|  |  | Hip Hop | Alternative | Post-rock | Total |
| Age level | Middle School | 30 | 16 | 4 | 50 |
| High School | 20 | 24 | 6 | 50 |
| College | 16 | 20 | 14 | 50 |
|  | Total | 66 | 60 | 24 | 150 |

(a) In a table like the one below, provide the appropriate conditional distribution for comparing the music-listening preferences of the three age levels.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Music Genre | | |
|  |  | Hip Hop | Alternative | Post-rock |
| Age level | Middle School |  |  |  |
| High School |  |  |  |
| College |  |  |  |

(b) Make a graph that illustrates these conditional distributions effectively, and use the table from (a) and your graph to describe the relationship between age level and preferred music genre.

(c) Perform the appropriate statistical test to determine if there is a difference in the music preference of these three age groups.

(d) If you chose a test for homogeneity in part (c), explain how the data could have been obtained to make a test for independence/association appropriate. If you chose a test for independence/association, explain how the data could have been obtained to make a test for homogeneity appropriate.

4. Does the season in which you are born affect the probability that you will be allergic to dust mites? Recent research suggests this might be true. Below are the birth seasons of 500 randomly selected people who are allergic to house dust mites, along with the expected proportion of people born in each season, based on data from the entire population. Do these data provide evidence that people who suffer from this allergy have a different birth season distribution? Support your conclusion with the appropriate statistical test.

|  |  |  |
| --- | --- | --- |
| Birth Season | Number of allergy  sufferers | Proportion of births  in general population |
| Winter | 117 | 0.30 |
| Spring | 105 | 0.22 |
| Summer | 145 | 0.24 |
| Fall | 133 | 0.24 |