**#1** Consider the following data (counts) for the relationship between number of ear piercings and whether a student has a tattoo for 1014 Penn State male students.

|  |  |  |  |
| --- | --- | --- | --- |
| **Ear Pierces** | **No tattoo** | **Tattoo** | **Total** |
| **0** | 699 | 68 | 767 |
| **1** | 87 | 27 | 114 |
| **2 or more** | 87 | 46 | 133 |
| **Total** | 873 | 141 | 1014 |

a. Use the counts given above to calculate **column percentages** for this table. Fill in the following table with those column percentages.

|  |  |  |
| --- | --- | --- |
| **Ear Pierces** | **No tattoo** | **Tattoo** |
| **0** | 0.800687285 | 0.482269504 |
| **1** | 0.099656357 | 0.191489362 |
| **2 or more** | 0.099656357 | 0.326241135 |
| **Total** |  |  |

b. Explain why the answer to part a of this activity suggests that there may be a relationship between number of ear piercings and whether a student has a tattoo.

**Those with tattoos are far more likely to have piercings (~52% have one or more) compared to those without tattoos (only ~20% have one or more)**

c. Here is some computer output for a (chi-square) test of the relationship between these two variables.

Pearson Chi-Square = 72.894, DF = 2, P-Value = 0.000

Write the null and alternative hypotheses that are being tested by this test

**H\_0: There is no relationship between having tattoos and having ear piercings**

**H\_A: There is a relationship between having tattoos and having ear piercings**

On the basis of the test, what are we able to conclude?

**Because p<.05, we reject the null and accept the alternative. We can conclude that there is a relationship between having tattoos and having ear piercings**

d. The result in part c was “statistically significant.” Explain what that means.

**The observed relationship was unlikely to have been produced by chance alone. We think this represents a real relationship in the population.**

**#2** Use the dataset ucdavis2.sav that can be found in the Lab3 folder of the course website. The data are from an introductory statistics class at UC Davis.

**a.** The variable ***Cheat*** gives responses to the question “Have you ever cheated on a significant other?” Use **Analyze>Descriptive Statistics>Crosstabs** to fill in the following two-way table with counts for the relationship between ***Sex*** and ***Cheat***. In the Cells dialog box, ask for row percentages as well. Be sure to disregard the missing values in the first column and their effects on the totals.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cheated on SO?** | |  |
| **Gender** | No | Yes | Total |
| Female | 132 | 18 | 150 |
| Male | 84 | 2 | 86 |
| Total | 216 | 20 | 236 |

Also fill in this two-way table with row percentages.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cheated on SO?** | |  |
| **Gender** | No | Yes | Total |
| Female | 88% | 12% | 100% |
| Male | 97.7% | 2.3% | 100% |
| Total | 91.5% | 8.5% | 100% |

**b.** Write a sentence that interprets the row percentage given in the “Female row and “No column” in part a.

**Of all females, 88% say they have never cheated on a significant other**

**c.** Explain whether you think that the table of percents found in part a is evidence of a relationship between the two variables in this activity or not.

**Yes – females are almost 6x as likely to say they have cheated on a significant other.**

**d.** Now go back into the Crosstabs menu, and in the Statistics dialog box, ask for Chi-square.

What is the *p*-value at the end of the “Pearson chi-square” line in the results? SPSS labels this as “Asymp. Sig. (2-sided)” **0.015**

Is the observed result “statistically significant?” Explain why or why not.

**Yes – p<.05**

**e.** Write a sentence than gives a conclusion about the two variables in this activity (gender and whether student has cheated on significant other.).

**The two variables are significantly related, so we can say females are significantly more likely to cheat as compared to males.**

**f.** Refer back to the chi-square test done in part d. Write the null and alternative hypotheses being tested by that chi-square test. Make your answer specific to the variables in this example.

**H\_0: There is no relationship between gender and having cheated on a SO**

**H\_A: There is a relationship between gender and having cheated on a SO**

**g.** You might notice that the footnote on the Chi-square value indicates that the assumptions of the test might not be met. This is an idiosyncrasy caused by those three missing values. If you sort the data by Cheat and find and delete the three missing values, you will see that the result does not change very much. Do this and report the new Chi-square p-value.

**.01**

**#3** Continue to use the ucdavis2.sav dataset. The variable ***Hand*** records handedness.

a. Use **Stat>Tables> Crosstabulation and Chi-square** to fill in the following table with counts *and* percentages by gender.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Handedness** | | |
| **Gender** | Left | Right | Total |
| Female | 12  7.8% | 141  92.2% | 153 |
| Male | 8  9.3% | 78  90.7% | 86 |

**b.** On the basis of percentages that you found for part a, describe the relationship between the two variables in this activity.

**Though males are slightly more likely to report being left-handed, it does not seem like a very strong relationship between these two variables.**

**c.** Do a chi-square test for statistical significance of the observed relationship.

(i) Give a *p*-value for the Pearson version of the test. **.696**

(ii) Explain whether the observed relationship is statistically significant.

**No – p>>.05, so what we observe is quite likely due to chance and does not represent a real relationship in the population.**

**#4** Continue to use the ucdavis2.sav data. The variable ***Seat*** gives student responses to the question “Where do you normally sit in the classroom? **B**ack of the room, **M**iddle of the room, or **F**ront of the room.” Analyze the relationship between ***Gender*** (female, male) and ***Seat***. Write a paragraph that summarizes your analysis. In your summary compare relevant conditional percentages to describe the relationship and report the result of a chi-square test of statistical significance. State a clear conclusion for the test of significance.

**When analyzing the relationship between gender and where students generally sit in the classroom, we find that both genders have the majority of respondents reporting that they sit in the middle of the classroom. For the minority who either usually sit in the front or back, females are almost twice as likely to sit in the front compared to the back, while for males they are almost twice as likely to sit in the back compared to the front. The Chi^2 test for this relationship has a p-value of .031, so we can conclude that this observed relationship represents a real phenomenon in the population, and the genders have a significantly different proportion of students sitting in the three locations in general.**

**#5** Use the Spring 2006 Survey dataset (SPR06FINAL.sav) in Lab3 folder of the course website. It gives data for a few questions asked in the Spring 2006 PSU Stat 200 survey. The variable ***FavMusic*** gives responses to the question, “What is your favorite type of music (Rock, Rap/HipHop, Pop, Country, Other)?” In this activity, we’ll look how this variable relates to gender.

**a.** For each sex, determine the sample percentages in each category of favorite music. (Percentages should add to 100% within each sex.) Give those percentages as the answer to this part.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Country** | **Other** | **Pop** | **Rap/Hip Hop** | **Rock** |
| **Female** | **14.5%** | **20.6%** | **20.6%** | **18.7%** | **25.6%** |
| **Male** | **8.3%** | **18.7%** | **3.6%** | **26.4%** | **43%** |

**b.** Using your answer to part a as the basis, write a brief description of how males and females differ with regard to their favorite type of music.

**Males are far more likely than females to cite Rock as their favorite, while females are far more likely than males to cite Country or Pop as their favorite music.**

**c.** Do a chi-square test of the relationship between gender and favorite type of music.

What is the *p*-value for the test? **0.000**

State a conclusion for the chi-square test. **We reject H­\_0 and conclude that there is a significant relationship between gender and which style of music is your favorite.**

**d.** The dataset also includes measurements (in centimeters) on left forearm lengths and left foot lengths of the students. Explain why we should not use a two-way table and a chi-square test to analyze the relationship between those two variables.

**Since forearm lengths are a continuous RV, we would not use a Chi^2 test, which is meant to compare two categorical variables.**