# CHAPTER 14

## Chi Square Tests

***Summary***

You were introduced to a new NHST technique in this chapter called chi square ($x^2$). Chi square is different from the previous NHST techniques you have learned because it is appropriate when the data consist of categories and their frequency counts, rather than scores (as has been the case for all the tests you have learned up until this chapter). Two kinds of chi square test were described in the chapter: $\chi^2$ ***test of independence*** and $\chi^2$ ***goodness-of-fit test***.

In $\chi^2$ *test of independence*, the null hypothesis is that the two variables are independent of one another (hence calling it the test of "independence"). Another way of stating this null hypothesis is to say that the two variables are *not* related. This will sound familiar to you, as it is the same null hypothesis you used for the Pearson product-moment correlation coefficient. Both tests ask the same question, but Pearson's correlation coefficient uses quantitative scores for data and $\chi^2$ uses frequency counts. For $\chi^2$ *test of independence*, a rejected null hypothesis means that the two variables are probably related; they are not independent. A retained null hypothesis means that the data support that the variables are not related; they are independent (though independence is *not* proven).

*Chi square* works by comparing the frequency counts in each category that actually occurred in the data (***observed frequencies; O***) to the frequency counts that would be expected in each category, if the null hypothesis were true (***expected frequencies; E***). In every case, $\Sigma O$ must equal $\Sigma E$.

For $\chi^2$ *test of independence*, expected frequencies are calculated using the data from the actual study and tell you what frequencies you'd expect if the two variable were *not* related. If the observed and expected frequencies are different enough, then the null hypothesis can be rejected. To determine if they are different enough, just like in previous NHST techniques, a table based on the $\chi^2$ family of distributions is consulted to find a critical value. Observed $\chi^2$ values that are larger than the critical value have a low probability when the null hypothesis is true and lead to rejection of the null hypothesis. Like the *t* and *F* distributions, the shape of the $\chi^2$ distribution is based on associated degrees of freedom. For $\chi^2$ *test of independence*, $df = (R\text{-}1)(C\text{-}1)$ where *R* stands for the number of rows in the ***contingency table*** and *C* stands for the number of columns. This is a bit different than other distributions you have learned about whose degrees of freedom are mostly determined by sample size. Like the *F* distribution, the $\chi^2$ distribution is positively skewed and becomes less skewed as *df* increases.

Just like other NHST techniques you have learned, the $\chi^2$ *test of independence* can be followed by an effect size index. The $x^2$ *test of independence* is used to determine *whether* there is a relationship between two variables and an effect size index indicates *how big* the relationship is. In this chapter, you were introduced to two effect size indices for the $\chi^2$ *test of independence*: *odds ratios* and $\Phi$ (in statistics, use the British pronunciation of this Greek letter, "fee"). *Odds ratios* are a measure of the likelihood of an event calculated by dividing one odd into another. Following a $\chi^2$ *test of independence*, they allow statements like "men are three times as likely as women to volunteer to participate in a study about risk-taking". Phi works only for $2 \times 2$ $\chi^2$ *test of independence*. Like Pearson's product-moment correlation coefficient, values of $\Phi$ of 0.10, 0.30, and 0.50 are considered small, medium, and large respectively. There are effect size indices that work for $\chi^2$ *test of independence* with contingency tables larger than $2 \times 2$, one example is *Cramer's* $\Phi$.

The $\chi^2$ *goodness-of-fit test* is similar to the $\chi^2$ *test of independence* in some ways and differs in others. Both use the logic of NHST and frequency counts rather than scores as data. They differ in the questions they address, the null hypothesis they test, how to determine expected frequencies, and effect size indices. The $\chi^2$ *goodness-of-fit test* examines whether the sample data fit a particular theory. The null hypothesis is that the data *fit* the theory. The alternative is that the data *do not fit* the theory. If the null hypothesis is rejected, conclude that the data are not consistent with the theory. If the null hypothesis is retained, conclude that the data are consistent with the theory and therefore support (but do not prove) the theory. For $\chi^2$ *goodness-of-fit test*, expected frequencies are determined from the theory rather than the actual collected data. Degrees of freedom for the $\chi^2$ *goodness-of-fit test* are equal to the number of categories minus one. There are no effect size indices for $\chi^2$ *goodness-of-fit test*.

Chi square tests are appropriate if 1) the measurement of a person or event places it into a category (frequency data not scores), 2) the measurement of one person or event does not influence the measurement of another person or event (the measurements are independent), and 3) samples are representative of the populations that the conclusions apply to.

Statisticians were once concerned that small expected frequencies would lead to increased chances of Type I errors. This does not seem to be the case, particularly if sample size is greater than 20. However, if one or more expected frequencies are small (5 or less), the probabilities given by the chi square distribution may lead to increased chances of *Type II errors*. The best solution is to gather sufficient data to avoid this problem. A second solution is to combine related categories, which results in fewer categories and larger expected frequencies.

***Multiple-Choice Questions***

1. You should choose a $\chi^2$ test when your data are ________.

a. scores
b. frequency counts
c. either a. or b.
d. neither a. nor b.

2. The shape of the theoretical $\chi^2$ distribution is determined by _______.

a. the number of observations
b. the size of the expected frequencies
c. the number of categories of events
d. all of the above

3. For $\chi^2$, the sum of the expected frequencies must be equal to _______.

a. the sum of the observed frequencies
b. the *df*
c. (*R*-1)(*C*-1)
d. none of the above

4. A developmental psychologist developed a theory that predicted the proportion of children who would, in a period of stress, cling to their mother, behave aggressively toward their mother, or behave aggressively toward a younger sibling. The stress situation was set up, and the responses of 50 children with their mother and younger sibling present were recorded and compared to the values predicted by the theory. The appropriate $\chi^2$ test is _______.

a. $\chi^2$ *goodness-of-fit test* with 2 *df*
b. $\chi^2$ *goodness-of-fit test* with 49 *df*
c. $\chi^2$ *test of independence* with 2 *df*
d. $\chi^2$ *test of independence* 49 *df*

5. The null hypothesis for a $\chi^2$ *goodness-of-fit test* is that ________.

a. the data fit the theory
b. the data do not fit the theory
c. the data are normally distributed
d. the data have equal variances

6. A political science researcher is interested in whether political affiliation (democrat, republican, independent) is related to likelihood of donating (or not) to a political campaign. She collects data, does a $\chi^2$ *test of independence*, and rejects the null hypothesis. The proper conclusion is that _______.

a. political affiliation and donating to political campaigns are not related to each other
b. political affiliation and donating to political campaigns are related to each other
c. political affiliation and donating to political campaigns are independent of each other
d. knowing a person's political affiliation gives you no clue to his or her likelihood of donating to a political campaign

7. A $\chi^2$ *goodness-of-fit test* was used to evaluate a theory. The null hypothesis was rejected. The proper conclusion is that the theory is _______.

a. supported by the data
b. not supported by the data
c. either a. or b., depending on the *df*
d. theories must be evaluated with $\chi^2$ *test of independence*

8. To use the $\chi^2$ distribution with confidence, you must assume that the observations you make _______.

a. are normally distributed
b. have equal variances
c. are independent
d. all of the above

9. The formula for finding expected frequencies in a $\chi^2$ *test of independence* _________.

a. assumes that the categories are independent
b. uses predictions based on a theory
c. assumes that the categories of events are related in some way
d. none of the above

10. For a 2 × 2 $\chi^2$ *test of independence*, ________ is an effect size index that can be used, and _______ is a medium value of that statistic.

a. $\theta$, 0.50
b. $\theta$, 0.30
c. $\Phi$, 0.50
d. $\Phi$, 0.30

11. Suppose you were analyzing data from a 2 × 2 $\chi^2$ *test of independence* and you had one expected frequency that was very small ($N = 2$). According to your text, you are very likely to make a ________.

a. Type I error
b. Type II error
c. both a. and b.
d. neither a. nor b.

12. If you compare $\chi^2$ distributions for skewness, you'll find that as degrees of freedom increase the curves are still _______.

a. positively skewed but becoming less so
b. positively skewed and becoming more so
c. negatively skewed but becoming less so
d. negatively skewed and becoming more so

13. After the data are gathered, a statistician may decide to combine logically-related categories. The most likely reason for this is to ______.

a. hide mistakes
b. increase the size of some expected frequencies
c. increase the degrees of freedom
d. decrease the skew of the chi square distribution

14. Examining the magnitude of each category's contribution to the final value of an observed $\chi^2$ is important because ________.

a. the sum of these contributions should always be equal to zero in a properly worked $\chi^2$
b. each of these contributions must be evaluated separately to determine if the null will be retained
c. it helps determine which category(ies) are likely to have meaningful differences between expected and observed frequencies
d. it is the best way to determine if the null hypothesis should be rejected

15. A sociobiology theory predicted that the amount of help offered by three groups would depend on the degree of kinship to the helped group. The degrees of kinship were 15 percent, 10 percent, and 0 percent. When the frequency of help data were analyzed, the $\chi^2$ value was smaller than the tabled value for the appropriate degrees of freedom. The data _______.

a. support the sociobiological theory
b. do not support sociobiological theory
c. more information is necessary to determine if the data do or do not support the sociobiological theory
d. indicate that there is no relationship between kinship and helping

16. Quinn is interested in whether there is a relationship between gender identity and helping behavior. Quinn gets a lot of information from participants including gender identity. Participants are then put in a situation where they can help or not help someone else. Quinn collects the data, runs an analysis, and finds a chi square value smaller than the tabled value. Quinn's analysis supports the idea that gender identity and helping behavior are _______.

a. independent
b. not independent
c. either a. or b. depending whether the calculated value is significantly smaller than the tabled value
d. either a. or b. depending on the shape of the $\chi^2$ distribution.

17. The $\chi^2$ *goodness-of-fit test* is useful to help determine if a set of data ______.

a. are consistent with a theory
b. show a relationship between two variables
c. have significant differences between means
d. all of the above

18. Assume you work in an airport food court that is open seven nights a week and you are interested in determining if customers prefer eating there on a particular night, relative to the others. You would expect, by chance, that there would be an equal number of customers eating in the food court each night. You count the number of customers eating in the food court each night for a week. Which $\chi^2$ test is most appropriate to analyze your data?

a. $\chi^2$ *goodness-of-fit test*
b. $\chi^2$ *test of independence*
c. neither a. nor b., you should use a paired samples *t* test
d. neither a. nor b., you should use a one-factor repeated measures ANOVA

19. Imagine you did the study from the last question to determine if there is a preference to eat in an airport food court on any of the seven nights a week. How many degrees of freedom should you report?

a. 8
b. 7
c. 6
d. 12

20. Suppose your data in a 2 × 2 $\chi^2$ *test of independence* produce an $\Phi$ of 0.30, you should consider this a _______ effect size.

a. large
b. medium
c. small
d. none of the above, $\Phi$ is not a measure of effect size

***Short Answer***

1. How are the data for a $\chi^2$ analysis different from the data for a Pearson's product-moment correlation coefficient?
2. Distinguish between when to use a $\chi^2$ *test of independence* and when to use a $\chi^2$ *goodness-of-fit test*.

3. In an earlier question, Quinn was interested in whether there is a relationship between gender identity and helping behavior. Identify the following.
   a. the appropriate $\chi^2$ test for Quinn to use
   b. the null and the alternative for Quinn's analysis

4. Earlier you imagined you worked in an airport food court that is open seven nights a week and you were interested in determining if customers prefer eating there on a particular night relative to the others. You would expect, by chance, that there would be an equal number of customers eating in the food court each night. In the earlier multiple choice problem, you decided to do a $\chi^2$ *goodness-of-fit test*. Now state the null and alternative for your study.

5. Based on all her prior work, an environmental biologist developed a theory that the kind of coral she studied would survive equally well whether their water had 3.25%, 3.35%, 3.45%, or 3.55% salt. She collected data and performed the appropriate $\chi^2$ analysis. She rejected the null and found that the coral were more likely to survive in the 3.45% and 3.55% salinity water. She followed up her analysis with $\Phi$ and determined that the effect size was large.
   a. What is wrong with the study description?
   b. Write a study description using similar variables that does not have the same problem.

***Problems***

1. Meera is an education major interested in variables that predict skipping college class. She counts the number of students skipping class on a single day in four classes at the same time at her University. She categorizes the students who skipped class as either taking the skipped class in their major or not. She also collects and categorizes the students' GPA's as low, medium, or high. Meera wants to know if GPA is related to skipping class in majors and nonmajors. Her data can be found below.

| | GPA | | |
|---|---|---|---|
| | Low | Medium | High |
| Majors | 21 | 25 | 31 |
| Non Majors | 35 | 21 | 14 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze her data with the appropriate test, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$.

2. A team of researchers is trying to choose the best incentive to use to entice participants to be in their study. They cannot agree on which incentive will be most preferred. So as any good team of researchers should, they collected some data. They asked 500 people which of five incentives they would rather receive for participating in a study that would require them to keep two days of food and activity diaries, complete 3 hours of questionnaires and an hour of computerized memory tasks, and abstain from caffeine for the two days prior to the study and the two study days. Below find how many of the 500 individuals preferred each incentive.

| | Incentives | | | | |
|---|---|---|---|---|---|
| | Guaranteed $25 Amazon.com gift card | Guaranteed $25 gift card to local popular restaurant | 1 in 500 chance to win $5000 | 1 in 100 chance to win $1000 | 1 in 25 chance to win $200 |
| # of participants selecting each incentive | 107 | 79 | 95 | 125 | 94 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze the research team's data with the appropriate test to determine which if any of the five incentives is/are preferred by potential participants, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$.

3. In 2015, staff members of the American Psychological Association's (APA's) Center for Workforce Studies published a report indicating that as of 2013, for every male active psychologist, there were 2.1 female active psychologists in the workforce. A college professor had noticed that she had far more female students than male students and wondered if the sex makeup for her undergraduate students matched the findings from the APA's report. She asked 310 of her students to self-identify their sex. If the students matched the report, she would expect 210 females and 100 males in her student sample. Her data follow.

| | Sex | |
|---|---|---|
| | Male | Female |
| Psychology Majors | 77 | 233 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze the professor's data with the appropriate test, calculate an effect size if you know an appropriate one, and write an interpretation. Let α = .05.

4. Christopher Olivola (2018) investigated the "sunk-cost fallacy." The sunk-cost fallacy occurs when we pursue an inferior alternative in a decision-making process because we have invested something (money, time, emotional resources, etc.) in that alternative in the past, instead of giving up that investment and making a wise choice. For example, the sunk-cost fallacy would predict people would keep a poorly performing stock, instead of selling it and getting a different stock, because money had already been spent on the poorly performing stock and selling would mean losing the investment they had made. Many studies have shown that one's own investment leads to an illogical decision because of the sunk-cost fallacy. Olivola had his participants imagine a scenario in which they had a non-refundable front-row ticket to a basketball game, but terrible weather on game day meant travel would be very treacherous and time-consuming. Participants were told to imagine that a) they had paid $200 for the ticket (high sunk-cost) or b) they had obtained the ticket for free (low sunk-cost). Each participant then decided if they would go to the game despite the weather or stay home and safely watch the game on television. The data below (retrieved from https://osf.io/nzp8t/ on July 23, 2018) are from Olivola's paid participants from Amazon's Mechanical Turk. You can see the number of participants in each category who chose the unwise option (to go ahead to the game in the treacherous weather) and the number choosing to let the investment go and stay home safely watching the game on television. For this question, do an analysis to determine if Olivola *replicated* the common observation of sunk-cost decisions. Is there a relationship between level of one's own investment and unwise choices?

| | Sunk-cost | |
|---|---|---|
| | Low (ticket was free) | High (ticket cost $200) |
| Chose wisely (stayed safely at home but lost investment) | 106 | 68 |
| Chose unwisely (went to game in treacherous weather) | 50 | 80 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze Olivola's data with the appropriate test, calculate an *odd's ratio* effect size if it is appropriate, and write an interpretation. Let α = .05.

5. Olivola (2018) also wanted to know if people will demonstrate the sunk cost fallacy when it is *someone else's* investment, *rather than their own*, that will be lost. In this part of the study, participants were told to imagine that a) a friend had paid $200 and given them the ticket (high *other* sunk-cost) or b) a friend had obtained the ticket for free and given it to them (low *other* sunk-cost). Now analyze Olivola's (2018) data to determine if participants are more likely to choose unwisely when *someone else* has invested in an option. Is there a relationship between level of investment and unwise choices when the investment was made by *someone else*?

| *OTHER* | Sunk-cost | |
|---|---|---|
| | Low (ticket was free) | High (ticket cost $200) |
| Chose wisely (stayed safely at home but lost investment) | 100 | 101 |
| Chose unwisely (went to game in treacherous weather) | 35 | 59 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze Olivola's data with the appropriate test, calculate a $\Phi$ effect size if it is appropriate, and write an interpretation. Let $\alpha = .05$.

6. Making sure that animals kept in zoos and other manmade environments have sufficient enrichment to maintain cognitive welfare is important. Hanna et al. (2017) examined a killer whale's response to visual media. In other words, did she like to watch television? And if she did, what did she like to watch? Relative to a control television that was turned off, they found that their killer whale spent seven times longer watching dolphin (prey of killer whales) documentaries, six times longer watching whale documentaries, and four times longer watching elephant documentaries. This makes for a ratio of 7(prey): 6(self): 4(unrelated animal): 1(control). A *dolphin* trainer wonders if her dolphins will have similar preferences to the killer whale in the Hanna et al. (2017) study. Over time, she releases 90 dolphins into a viewing tank that had four televisions in it. One of the televisions played videos of dolphin prey (fish, squid, shrimps, jellyfish and octopuses), one of the televisions played dolphin videos, one played elephant videos, and the control television was turned off. Below, find how many of the 90 dolphins preferred each of the four televisions.

| | Visual Media | | | |
|---|---|---|---|---|
| | Prey | Self | Elephants | Control |
| Dolphins | 32 | 32 | 22 | 4 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze the dolphin trainer's data with the appropriate test to determine if her dolphins had similar visual media preferences to the killer whale, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$.

7. In Chapter 11, you were introduced to a study about sharing. Here's a description of the study: *Researchers in Poland (Gasiorowska, Chaplin, Zaleskiewicz, Wygrab, & Vohs, 2016) were interested in whether children primed to think about money shared less than children primed to think about other things. These researchers had 3-6 year old Polish children sort either coins, buttons, or candy for 10 minutes (this technique is known as ''priming''). Afterward, the children were shown six Disney character stickers (Note: If you do not know much about children, Disney character stickers are highly coveted items). They were allowed to take as many stickers as they wanted. Next, they were given the opportunity to give some, all, or none of their stickers away to a child who they were told did not get to participate in the study.* In Chapter 11, you used ANOVA to determine if the percentage of stickers children gave were related to what they were primed to think about in the sorting task. The researchers also examined if there was a relationship between what children were primed to think about and what they called ''extreme selfishness,'' which they defined as giving none of their stickers to the other child. The data below are the number of children in each group who donated nothing (demonstrating extreme selfishness) or something to another child.

| | Sorting Group | | |
|---|---|---|---|
| | Money | Buttons | Candy |
| Donated nothing | 15 | 4 | 6 |
| Donated something | 27 | 38 | 36 |

a. Choose the appropriate $\chi^2$
b. State the null hypothesis
c. Analyze their data with the appropriate test, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$.

**ANSWERS**

***Multiple-Choice Questions***

1. b
   **Explanation:** Chi square is different from other tests you have learned so far because the data are frequency counts rather than scores.
2. c
   **Explanation:** Chi square is different from other tests you have learned so far because the degrees of freedom which determine shape of the distribution are not dependent on the number of observations (sample size), but on the number of categories the data are divided into.
3. a
   **Explanation:** Whenever a $\chi^2$ problem has been worked correctly the $\Sigma O$ must equal $\Sigma E$. Whenever you are working a problem, it is a good idea once you have determined the observed and expected frequencies to stop and check that you have not made a mistake by making sure this is true.
4. a
   **Explanation:** Here the developmental psychologist has a theory she wishes to test about the number of children that will behave in each of three ways. Because her predictions are based on a theory, it is a $\chi^2$ *goodness-of-fit test*. For $\chi^2$ *goodness-of-fit test*, *df* = # of categories – 1; here, 3 – 1 = 2.
5. a
6. b
   **Explanation:** Choices a., c., and d., are all variations of the null hypothesis for this study. Remember, the null hypothesis in $\chi^2$ *test of independence* is that the variables are independent.
7. b
   **Explanation:** Remember for $\chi^2$ *goodness-of-fit test*, the null hypothesis is that the data *fit* the theory. If you reject the null, that suggests the data *do not fit* the theory.
8. c
   **Explanation:** Chi square tests are appropriate if 1) data are frequency counts, 2) the measurements are independent, and 3) samples are representative of the populations that the conclusions apply to. There are no assumptions about normality or homogeneity of variance as were seen for *t* tests and ANOVA.
9. a
   **Explanation:** In all NHST techniques, including $\chi^2$, you begin with the assumption that the null hypothesis is true.
10. d
    **Explanation:** Although they may look somewhat similar, the symbol in a. and b. is theta, not phi.
11. b
    **Explanation:** Statisticians were once concerned that small expected frequencies would lead to increased chances of Type I errors. This does not seem to be the case, particularly if sample size is greater than 20. However, if one or more expected frequencies are small (5 or less), the probabilities given by the chi square distribution may lead to increased chances of *Type II errors*.
12. a
13. b
14. c
    **Explanation:** If a category has a large $\frac{(O-E)^2}{E}$ value associated with it, that is a strong clue that there are meaningful differences between the observed and expected values for that category. The reason you want to know which category or categories had meaningful differences between the expected and observed frequencies is that you will use this to describe what happened in your interpretation of the problem.
15. a
    **Explanation:** Because the chi square value was smaller than the tabled value you should retain the null hypothesis. This is a $\chi^2$ *goodness-of-fit test*. The null hypothesis is that the data fit the theory. Because the null is retained, the data supports the theory used in the problem.
16. a
    **Explanation:** Because the chi square value was smaller than the tabled value you should retain the null hypothesis. This is a $\chi^2$ *test of independence*. The null hypothesis is that the two variables are independent (not related). Because the null is retained, the analysis supports the idea that gender identity and helping behavior are independent.
17. a
18. a
    **Explanation:** Here, your 'theory' is that the same number of customers will eat in the food court each night.
19. c
    **Explanation:** Because this is a $\chi^2$ *goodness-of-fit test* the degrees of freedom are # of categories – 1; here, because there are 7 nights in a week, 7 – 1 = 6.
20. b
    **Explanation:** Like Pearson's product-moment correlation coefficient, $\Phi$ values of 0.10, 0.30, and 0.50 are considered small, medium, and large respectively.

***Short Answer***

1. $\chi^2$ analysis requires that data are categories and their frequency counts, rather than quantitative scores which are required by Pearson's product-moment correlation coefficient.

2. Use a $\chi^2$ *test of independence* to determine if two variables are related or independent. Use a $\chi^2$ *goodness-of-fit test* to determine if a set of data supports or refutes a theory.

3. a. $\chi^2$ *test of independence*
   b. $H_0$: There is no relationship between gender identity and helping behavior, OR gender identity and helping behavior are independent.
   $H_1$: There is a relationship between gender identity and helping behavior, OR gender identity and helping behavior are not independent.

4. $H_0$: The data fit the theory; there will be equal numbers of customers eating in the food court each night.
   $H_1$: The data do not fit the theory; there will be unequal numbers of customers eating in the food court each night.

5. a. The environmental biologist is doing a $\chi^2$ *goodness-of-fit test*. There are no effect size indices for this test. $\Phi$ only works for a 2 × 2 $\chi^2$ *test of independence.*
   b. You could write many versions of a study description that work here as long as you describe a 2 × 2 $\chi^2$ *test of independence.* (Technically, you could write the study above and leave off the $\Phi$ , but that wouldn't give you practice thinking about the $\chi^2$ *test of independence*.) Here is an example of a study that would use $\chi^2$ *test of independence* with these variables: An environmental biologist wonders if there is a relationship between the salinity of water (3.45% or 3.25%) and the survival of a coral that she studies (survive or not). She collected data, performed the appropriate $\chi^2$ analysis. She rejected the null and found that coral were more likely to survive in the water with 3.45% salinity and more likely to perish in the water with 3.25% salinity. She followed up her analysis with $\Phi$ and determined that the effect size was large.

***Problems***

1. a. $\chi^2$ *test of independence*
   b. $H_0$*:* There is no relationship between major and GPA with regard to skipping class.
   c. Because $\chi^2(2)_{.05} = 5.991$ and $\chi^2(2) = 9.96$, the null hypothesis can be rejected, the two variables were related, $p < .05$. Among those with the lowest GPAs, nonmajors were more likely to skip class (62.50% of the low GPA students who skipped class were nonmajors). However, among those with high GPAs, majors were more likely to skip class (68.89% of the high GPA students who skipped class were majors).

   APA format: There was a significant relationship between major status and GPA with regard to skipping class, $\chi^2(2) = 9.96$, $p < .05$. Among those with the lowest GPAs, nonmajors were more likely to skip class (62.50%) than majors (37.5%). However, among those with high GPAs, majors were more likely to skip class (68.89%) than nonmajors (31.11%).

   **Explanation:** Because Meera wants to know about a relationship between majoring and GPA with regard to skipping class, you should choose a $\chi^2$ *test of independence*.
   The null hypothesis is: There is no relationship between majoring and GPA with regard to skipping class.

Set up a contingency table to find the row and column totals you need.

| | GPA | | | |
|---|---|---|---|---|
| | Low | Medium | High | Σ |
| Majors | 21 | 25 | 31 | 77 |
| Nonmajors | 35 | 21 | 14 | 70 |
| Σ | 56 | 46 | 45 | 147 |

Now calculate the expected frequencies for each cell using this formula.

$$E = \frac{(RowTotal)(ColumnTotal)}{N_{tot}}$$

| | GPA | | |
|---|---|---|---|
| | Low | Medium | High |
| Majors | $\frac{(77)(56)}{147} = 29.33$ | $\frac{(77)(46)}{147} = 24.10$ | $\frac{(77)(45)}{147} = 23.57$ |
| Nonmajors | $\frac{(70)(56)}{147} = 26.67$ | $\frac{(70)(46)}{147} = 21.90$ | $\frac{(70)(45)}{147} = 21.43$ |

Now calculate the $\chi^2$ using this formula, set up in a table like below

$$\chi^2 = \sum \left[\frac{(O - E)^2}{E}\right]$$

| *O* | *E* | *O-E* | $(O-E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 21 | 29.33 | -8.33 | 69.39 | 2.37 |
| 25 | 24.10 | 0.90 | 0.81 | 0.03 |
| 31 | 23.57 | 7.43 | 55.20 | 2.34 |
| 35 | 26.67 | 8.33 | 69.39 | 2.60 |
| 21 | 21.90 | -0.90 | 0.81 | 0.04 |
| 14 | 21.43 | -7.43 | 55.20 | 2.58 |
| $\Sigma O = 147$ | $\Sigma E = 147$ | | | $\chi^2 = 9.96$ |

Now is a good time to stop and make sure ΣO = ΣE.

To find the critical value, first calculate degrees of freedom, then refer to Table E.
$df = (R\text{-}1)(C\text{-}1) = (2\text{-}1)(3\text{-}1) = (1)(2) = 2$
$\chi^2(2)_{.05} = 5.991$

Because the observed $\chi^2(2) = 9.96$, which is greater than $\chi^2(2)_{.05} = 5.991$, reject the null hypothesis and conclude there is a significant relationship between majoring and GPA with regard to skipping class.

Examine the last column of your $\chi^2$ calculation table and determine which cells are contributing the most to the significant $\chi^2$.
Majors with low GPAs are less likely to skip class than expected if the null were true.
Majors with high GPAs are more likely to skip class than expected if the null were true.
Nonmajors with low GPAs are more likely to skip class than expected if the null were true.
Nonmajors with high GPAs are less likely to skip class than expected if the null were true.

Condense these four sentences into something clear and succinct for your reader.
Because this is not a 2 × 2 $\chi^2$ *test of independence*, the *odds ratio* formulas and *Φ* you learned in this chapter will not work to test effect size.

2. a. $\chi^2$ *goodness-of-fit test*

b. The five incentives are all equally likely to be a participant's most preferred incentive.

c. Because $\chi^2(4)_{.05} = 9.488$ and $\chi^2(4) = 11.76$, the null hypothesis is rejected; the incentives were not all equally preferred. There was a significant difference in preference for the incentives, $p < .05$. Fewer people preferred a $25 gift card to a local popular restaurant and more people preferred a 1 in 100 chance to win $1000 than would be expected if all of the incentives were equally preferred. The researchers should consider the 1 in 100 chance to win $1000 and not the $25 restaurant gift cards as their incentive. They should note that the other choices were all preferred by approximately 1/5th of their potential participant pool. They might think about offering more than one incentive to attract the most participants.

APA format: There was a significant difference in preference among the five incentives, $\chi^2(4) = 11.76$, $p < .05$. Fewer people preferred a $25 gift card to a local popular restaurant (15.80%) and more people preferred a 1 in 100 chance to win $1000 (25%) than would be expected if all of the incentives were equally preferred (20%). The researchers should consider the 1 in 100 chance to win $1000 and not the $25 restaurant gift cards as their incentive. They should note that the other choices were all preferred by approximately 20% of their potential participant pool. They might think about offering more than one incentive to attract the most participants.

**Explanation:**

| | Incentives | | | | |
|---|---|---|---|---|---|
| | Guaranteed $25 Amazon.com gift card | Guaranteed $25 gift card to local popular restaurant | 1 in 500 chance to win $5000 | 1 in 100 chance to win $1000 | 1 in 25 chance to win $200 |
| # of participants selecting each incentive | 107 | 79 | 95 | 125 | 94 |
| Expected Frequencies | 100 | 100 | 100 | 100 | 100 |

| *O* | *E* | *O-E* | *(O-E)*$^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 107 | 100 | 7 | 49 | 0.49 |
| 79 | 100 | -21 | 441 | 4.41 |
| 95 | 100 | -5 | 25 | 0.25 |
| 125 | 100 | 25 | 625 | 6.25 |
| 94 | 100 | -6 | 36 | 0.36 |
| $\Sigma O = 500$ | $\Sigma E = 500$ | | | $\chi^2 = 11.76$ |

$df$ = # of categories - 1 = 5 - 1 = 4

$\chi^2(4)_{.05} = 9.488$

$\chi^2 = 11.76 > \chi^2_{.05} = 9.488$, reject the null hypothesis and conclude the incentives are not equally preferred.

Examine the last column of your $\chi^2$ calculation table and determine which cells are contributing the most to the significant $\chi^2$. Here fewer people prefer the $25 gift card to a local popular restaurant and more people prefer a 1 in 100 chance to win $1000 than predicted by the null. The other three options do not contribute much to the $\chi^2$ value.

Because this is a $\chi^2$ *goodness-of-fit test, there is not an appropriate measure of effect size.*

3. a. $\chi^2$ *goodness-of-fit test*

   b. The current student data will fit the 2013 data, which has a 2.1 female to 1 male ratio.

   c. Because $\chi^2(1)_{.05} = 3.841$ and $\chi^2(1) = 7.81$, the null hypothesis can be rejected; current students did not have the same sex makeup as the 2013 workforce reported in the APA report, $p < .05$. There were fewer men and more women in the undergraduate sample than the 2013 workforce ratio of 2.1:1 predicts. This suggests that when the professor's students get to the workforce there will be even fewer men compared to women.

APA format: The students' sex makeup was significantly different from the 2013 workforce reported in the APA report, $\chi^2(1) = 7.81$, $p < .05$. There were fewer men and more women than the 2013 workforce ratio of 2.1:1 predicted. This suggests that when the professor's students get to the workforce there will be even fewer men compared to women.

If you weren't sure how we got the expected frequencies for this problem, the report said that there were 2.1 females for every 1 male in the workforce. Remember from your study of probability you can calculate the probability of something happening by using this formula $p(A) = \frac{\#\ of\ events\ classifiable\ as\ A}{total\ \#\ of\ events}$ so here the probability of a female is $p(female) = \frac{2.1}{3.1} = .677$ (that denominator is 2.1 females + 1 male) and the probability of a male is $p(male) = \frac{1}{3.1} = .323$. Now you can use those probabilities to determine the number you expect from any sample size to be male or female by multiplying the probability by the sample size. So here, 310 participants × .677 = 210 females; and 310 participants × .323 = 100 males.

**Explanation:**

| | Sex | |
|---|---|---|
| | Male | Female |
| Psychology Majors | 77 | 233 |
| Expected Frequencies | 100 | 210 |

| *O* | *E* | *O-E* | *(O-E)²* | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 77 | 100 | -23 | 529 | 5.29 |
| 233 | 210 | 23 | 529 | 2.52 |
| $\Sigma O = 310$ | $\Sigma E = 310$ | | | $\chi^2 = 7.81$ |

$df$ = # of categories - 1 = 2 - 1 = 1

$\chi^2(1)_{.05} = 3.841$

$\chi^2 = 7.81 > \chi^2_{.05} = 3.841$, reject the null hypothesis and conclude the current student data do not fit the 2013 data, 2.1 female to 1 male ratio.

Look at the observed frequencies and expected frequencies and see why they didn't fit. There are fewer men and more women than the theory of 2.1:1 predicted.

Because this is a $\chi^2$ *goodness-of-fit test, there is not an appropriate measure of effect size.*

4. a. $\chi^2$ *test of independence*
   b. There is no relationship between level of one's own investment and unwise choices.
   c. Olivola replicated findings of previous research. Because $\chi^2(1)_{.05} = 3.841$ and $\chi^2(1) = 15.03$, the null hypothesis can be rejected; the two variables were related, $p < .05$. How much one had invested in a choice was significantly related to whether they will choose that option in the future, even if it is an unwise choice. People with high sunk-costs were more likely to choose unwisely and people with low sunk-costs were more likely to choose wisely than expected if there were no relationship between the two variables. Odds ratio analysis showed that people with high sunk-costs were 2.51 times more likely to choose unwisely (by selecting the bad choice they had invested in) compared to those with low sunk-costs.

   APA format: Olivola replicated findings of previous research. How much one has invested in a choice is significantly related to whether they will choose that option in the future, even if it is an unwise choice, $\chi^2(1) = 15.03$, $p < .05$. People with high sunk-costs were 2.51 times more likely ($OR = 2.51$) to choose unwisely (by selecting the bad choice they had invested in) compared to those with low sunk-costs.

**Explanation:**

| | Sunk-cost | | |
|---|---|---|---|
| | Low (ticket was free) | High (ticket cost $200) | Σ |
| Chose wisely (stayed safely at home but lost investment) | 106 | 68 | 174 |
| Chose unwisely (went to game in treacherous weather) | 50 | 80 | 130 |
| Σ | 156 | 148 | 304 |

| | Sunk-cost | |
|---|---|---|
| | Low | High |
| Chose wisely | $\frac{(174)(156)}{304} = 89.29$ | $\frac{(174)(148)}{304} = 84.71$ |
| Chose unwisely | $\frac{(130)(156)}{304} = 66.71$ | $\frac{(130)(148)}{304} = 63.29$ |

| *O* | *E* | *O-E* | $(O-E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 106 | 89.29 | 16.71 | 279.22 | 3.13 |
| 68 | 84.71 | -16.71 | 279.22 | 3.30 |
| 50 | 66.71 | -16.71 | 279.22 | 4.19 |
| 80 | 63.29 | 16.71 | 279.22 | 4.41 |
| $\Sigma O = 304$ | $\Sigma E = 304$ | | | $\chi^2 = 15.03$ |

$df = (R-1)(C-1) = (2-1)(2-1) = (1)(1) = 1$
$\chi^2(1)_{.05} = 3.841$
$\chi^2 = 15.03 > \chi^2_{.05} = 3.841$, reject the null hypothesis and conclude there is a significant relationship between level of one's own investment and unwise choices.

*odds ratio*
First, determine the odds of choosing unwisely when your own sunk-cost is high.
Odds = $\frac{number\ who\ do}{number\ who\ don't}$
Odds of choosing unwisely when own sunk-cost is high = $\frac{80}{68}$ = 1.18

Next, determine the odds of choosing unwisely when your own sunk-cost is low.

Odds of choosing unwisely when own sunk-cost is low = $\frac{50}{106}$ = 0.47

Now calculate the *odds ratio* that shows the likelihood of choosing unwisely when your own sunk-cost is high compared to when sunk-cost is low.

Odds ratio = $\frac{odds\ of\ Event\ A}{odds\ of\ Event\ B} = \frac{odds\ of\ choosing\ unwisely\ when\ sunk-cost\ is\ high}{odds\ of\ choosing\ unwisely\ when\ sunk-cost\ is\ low} = \frac{1.18}{0.47} = 2.51$

Conclude people with high sunk-costs are 2.51 times more likely to choose unwisely (by selecting the bad choice they had invested in) compared to those whose sunk-costs are low.

Note: you may have selected other odds and organizations of the odds ratio to examine (another way that works well is to examine the odds of making a *wise* choice based on sunk-cost level). Just make sure your technique accurately addresses the question Olivola had.

5. a. $\chi^2$ *test of independence*

b.There is no relationship between level of someone else's investment and one's own unwise choices.

c. Because $\chi^2(1)_{.05} = 3.841$ and $\chi^2(1) = 4.05$, the null hypothesis can be rejected; the two variables were related, $p < .05$. How much someone else has invested in a choice was significantly related to whether participants would choose that option in the future, even if it is an unwise choice. People who knew someone else had invested very little in an unwise option were less likely to choose that option and people who knew someone else had invested a lot in an unwise option were more likely to choose that option than would be expected if there were no relationship between the two variables. The effect was small ($\Phi = 0.12$).

APA format: Olivola extended previous research by examining the sunk-cost fallacy when it is *someone else's* investment that will be lost. There was a small, but significant, relationship between how much one believes someone else has invested in a choice option and whether they will choose that option in the future, even if it is an unwise choice, $\chi^2(1) = 4.05$, $p < .05$, $\Phi = 0.12$. People who knew someone else had invested very little in an unwise option were less likely to choose that option and people who knew someone else had invested a lot in an unwise option were more likely to choose that option than if there were no relationship.

**Explanation:**

| *OTHER* | Sunk-cost | | |
|---|---|---|---|
| | Low (ticket was free) | High (ticket cost $200) | Σ |
| Chose wisely (stayed safely at home but lost investment) | 100 | 101 | 201 |
| Chose unwisely (went to game in treacherous weather) | 35 | 59 | 94 |
| Σ | 135 | 160 | 295 |

| *OTHER* | Sunk-cost | |
|---|---|---|
| | Low | High |
| Chose wisely | $\frac{(201)(135)}{295} = 91.98$ | $\frac{(201)(160)}{295} = 109.02$ |
| Chose unwisely | $\frac{(94)(135)}{295} = 43.02$ | $\frac{(94)(160)}{295} = 50.98$ |

| *O* | *E* | *O-E* | *(O-E)*$^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 100 | 91.98 | 8.02 | 64.32 | 0.70 |
| 101 | 109.02 | -8.02 | 64.32 | 0.59 |
| 35 | 43.02 | -8.02 | 64.32 | 1.50 |
| 59 | 50.98 | 8.02 | 64.32 | 1.26 |
| $\Sigma O = 295$ | $\Sigma E = 295$ | | | $\chi^2 = 4.05$ |

$df = (R\text{-}1)(C\text{-}1) = (2\text{-}1)(2\text{-}1) = (1)(1) = 1$
$\chi^2(1)_{.05} = 3.841$

$\chi^2 = 4.05 > \chi^2_{.05} = 3.841$, reject the null hypothesis and conclude there is a significant relationship between level of someone else's investment and unwise choices.

*phi*

$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{4.05}{295}} = \sqrt{0.014} = 0.12$ small effect.

6. a. $\chi^2$ *goodness-of-fit test*
   b. The dolphins will have the same TV preferences as the killer whale.
   c. Analyze the dolphin trainer's data with the appropriate test to determine if her dolphins had similar visual media preferences to the killer whale, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$:
   Because $\chi^2(3)_{.05} = 7.815$ and $\chi^2(3) = 0.79$, the null hypothesis is retained; the dolphin data fit the theory based on the killer whale preferences. This suggests that the dolphins have the same preferences as the killer whale, $p > .05$. Like the killer whale, dolphins preferred television to no television and among the viewing choices they preferred videos of potential prey and other dolphins (self) to animals they do not likely recognize (elephants).

   APA format: The dolphin data did not differ significantly from the killer whale preferences, $\chi^2(3) = 0.79$, $p > .05$. This suggests that, like the killer whale, dolphins preferred television to no television and among the viewing choices they preferred videos of potential prey and other dolphins (self) to animals they do not likely recognize (elephants).

**Explanation:**

| | Visual Media | | | |
|---|---|---|---|---|
| | Prey | Self | Elephants | Control |
| Dolphins | 32 | 32 | 22 | 4 |
| Expected Frequencies | 35 | 30 | 20 | 5 |

If you weren't sure how to get the expected frequencies for this problem, the whale's preferences were 7:6:4:1. Some of you might have noticed that 7 + 6 + 4 + 1 = 18 and 18 divides evenly into 90 by 5. So multiplying each of the values in the odds by 5 will give you the expected frequencies. Alternately, remember from your study of probability, you can calculate the probability of something happening by using this formula $p(A) = \frac{\#\ of\ events\ classifiable\ as\ A}{total\ \#\ of\ events}$ so here the probability of a prey video being preferred is $p(prey) = \frac{7}{18} = .389$, now multiply that probability by the sample size 90 to get an expected value of 35. Repeat that process for each of the visual media types.

| $O$ | $E$ | $O\text{-}E$ | $(O\text{-}E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 32 | 35 | -3 | 9 | 0.26 |
| 32 | 30 | 2 | 4 | 0.13 |
| 22 | 20 | 2 | 4 | 0.20 |
| 4 | 5 | -1 | 1 | 0.20 |
| $\Sigma O = 90$ | $\Sigma E = 90$ | | | $\chi^2 = 0.79$ |

$df$ = # of categories - 1 = 4 - 1 = 3
$\chi^2(3)_{.05} = 7.815$
$\chi^2 = 0.79 < \chi^2_{.05} = 7.815$, retain the null hypothesis and conclude the dolphin data fits the theory from the killer whale, the dolphins have the same preferences as the killer whale.

Because this is a $\chi^2$ *goodness-of-fit test, there is not an appropriate measure of effect size.*

7. a. $\chi^2$ *test of independence*

b. There is no relationship between what children are primed to think about and extreme selfishness.

c. Analyze their data with the appropriate test, calculate an effect size if you know an appropriate one, and write an interpretation. Let $\alpha = .05$:

Because $\chi^2(2)_{.05} = 5.991$ and $\chi^2(2) = 10.28$, the null hypothesis can be rejected; the two variables were related. What children are primed to think about was significantly related to extreme selfishness, $p < .05$. Children primed to think about money were more likely to be extremely selfish and those primed with buttons were less likely to be extremely selfish than expected if the variables were independent. Children primed to think about money were more likely to be extremely selfish, 36% of them, compared to 10% of those primed with buttons and 14% of those primed with candy.

APA format: There was a significant relationship between what children were primed to think about and extreme selfishness, $\chi^2(2) = 10.28$, $p < .05$. Children primed to think about money were more likely to be extremely selfish, 36% of them, compared to 10% of those primed with buttons and 14% of those primed with candy.

| | Sorting Group | | | |
|---|---|---|---|---|
| | Money | Buttons | Candy | Σ |
| Donated nothing | 15 | 4 | 6 | 25 |
| Donated Something | 27 | 38 | 36 | 101 |
| Σ | 42 | 42 | 42 | 126 |

| | Sorting Group | | |
|---|---|---|---|
| | Money | Buttons | Candy |
| Donated nothing | $\frac{(25)(42)}{126} = 8.33$ | $\frac{(25)(42)}{126} = 8.33$ | $\frac{(25)(42)}{126} = 8.33$ |
| Donated something | $\frac{(101)(42)}{126} = 33.67$ | $\frac{(101)(42)}{126} = 33.67$ | $\frac{(101)(42)}{126} = 33.67$ |

| $O$ | $E$ | $O\text{-}E$ | $(O\text{-}E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 15 | 8.33 | 6.67 | 44.49 | 5.34 |
| 4 | 8.33 | -4.33 | 18.75 | 2.25 |
| 6 | 8.33 | -2.33 | 5.43 | 0.65 |
| 27 | 33.67 | -6.67 | 44.49 | 1.32 |
| 38 | 33.67 | 4.33 | 18.75 | 0.56 |
| 36 | 33.67 | 2.33 | 5.43 | 0.16 |
| $\Sigma O = 126$ | $\Sigma E = 126$ | | | $\chi^2 = 10.28$ |

$df = (R\text{-}1)(C\text{-}1) = (2\text{-}1)(3\text{-}1) = (1)(2) = 2$

$\chi^2(2)_{.05} = 5.991$

$\chi^2 = 10.28 > \chi^2_{.05} = 5.991$, reject the null hypothesis and conclude there is a significant relationship between extreme selfishness and what children are primed to think about.

Because this is not a 2 × 2 $\chi^2$ *test of independence*, the *odds ratio* and $\Phi$ you learned in this chapter will not work to test effect size.

***References***

American Psychological Association (2015). Demographics of the U.S. psychology workforce: Findings from the American Community Survey. Washington, DC: Author

Gasiorowska, A., Chaplin, L. N., Zaleskiewicz, T., Wygrab, S., & Vohs, K. (2016). Money cues increase agency and decrease prosociality among children: Early signs of market-mode behaviors. *Psychological Science, 27*(3), 331-344. doi: 10.1177/0956797615620378

Hanna, P., Brown, B., Winship, K., Cameron, D., Hill, H. M., & Kuczaj, S., II. (2017). A killer whale's (*Orcinus orca*) response to visual media. *International Journal of Comparative Psychology, 30*.

Olivola, C. Y. (2018). The interpersonal sunk-cost effect. *Psychological Science, 29*(7), 1072-1083. doi: 10.1177/0956797617752641