Chapter 11: FACTS ABOUT THE CHI-SQUARE DISTRIBUTION

|  |  |  |  |
| --- | --- | --- | --- |
| Exercise 1. | *If the number of degrees of freedom for a chi-square distribution is 25, what is the population mean and standard deviation?* | | |
| Solution | mean = 25 and standard deviation = 7.0711 | | |
| Exercise 2. | *If df > 90, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_. If df = 15, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.* | | |
| Solution | approximately normal, skewed right. | | |
| Exercise 3. | *When does the chi-square curve approximate a normal distribution?* | | |
| Solution | when the number of degrees of freedom is greater than 90 | | |
| Exercise 4. | *Where is μ located on a chi-square curve?* | | |
| Solution | just to the right of the peak | | |
| Exercise 5. | *Is it more likely the df is 90, 20 or two in the graph?*  *G:\Clients\Connexions\CONNEX120012_Statistics\07_Art\Chapter 11\CNX_Stats_C11_M03_item001.jpg*  Figure 11.13 | | |
| Solution | *df* = 2 | | |
| Exercise 6. | *Determine the appropriate test to be used. An archeologist is calculating the distribution of the frequency of the number of artifacts she finds in a dig site. Based on previous digs, the archeologist creates an expected distribution broken down by grid sections in the dig site. Once the site has been fully excavated, she compares the actual number of artifacts found in each grid section to see if her expectation was accurate.* | | |
| Solution | a goodness-of-fit test | | |
| Exercise 7. | *Determine the appropriate test to be used. An economist is deriving a model to predict outcomes on the stock market. He creates a list of expected points on the stock market index for the next two weeks. At the close of each day’s trading, he records the actual points on the index. He wants to see how well his model matched what actually happened.* | | |
| Solution | a goodness-of-fit test | | |
| Exercise 8. | *Determine the appropriate test to be used. A personal trainer is putting together a weight-lifting program for her clients. For a 90-day program, she expects each client to lift a specific maximum weight each week. As she goes along, she records the actual maximum weights her clients lifted. She wants to know how well her expectations met with what was observed.* | | |
| Solution | a goodness-of-fit test | | |
| Exercise 9. | *A teacher predicts that the distribution of grades on the final exam will be as in Table 11.27.*   |  |  | | --- | --- | | **Grade** | **Proportion** | | A | 0.25 | | B | 0.30 | | C | 0.35 | | D | 0.10 |   *Table 11.27*  *The actual distribution for a class of 20 is in Table 11.28.*   |  |  | | --- | --- | | Grade | Frequency | | A | 7 | | B | 7 | | C | 5 | | D | 1 |   *Table 11.28*  *df* = \_\_\_\_\_\_ | | |
| Solution | 3 | | |
| Exercise 10. | *A teacher predicts that the distribution of grades on the final exam will be in Table 11.27.*   |  |  | | --- | --- | | **Grade** | **Proportion** | | A | 0.25 | | B | 0.30 | | C | 0.35 | | D | 0.10 |   *Table 11.27*  *The actual distribution for a class of 20 is in Table 11.28.*   |  |  | | --- | --- | | Grade | Frequency | | A | 7 | | B | 7 | | C | 5 | | D | 1 |   *Table 11.28*  *State the null and alternative hypotheses.* | | |
| Solution | *H0* : The observed class grades fit the expected grades. *Ha* : the observed class grades do not fit the expected grades | | |
| Exercise 11. | *A teacher predicts that the distribution of grades on the final exam will be in Table 11.27.*   |  |  | | --- | --- | | **Grade** | **Proportion** | | A | 0.25 | | B | 0.30 | | C | 0.35 | | D | 0.10 |   *Table 11.27*  *The actual distribution for a class of 20 is in Table 11.28.*   |  |  | | --- | --- | | Grade | Frequency | | A | 7 | | B | 7 | | C | 5 | | D | 1 |   *Table 11.28*  *χ2 test statistic =\_\_\_\_\_\_* | | |
| Solution | 2.04 | | |
| Exercise 12. | *A teacher predicts that the distribution of grades on the final exam will be in Table 11.27.*   |  |  | | --- | --- | | **Grade** | **Proportion** | | A | 0.25 | | B | 0.30 | | C | 0.35 | | D | 0.10 |   *Table 11.27*  *The actual distribution for a class of 20 is in Table 11.28.*   |  |  | | --- | --- | | Grade | Frequency | | A | 7 | | B | 7 | | C | 5 | | D | 1 |   *Table 11.28*  *p-value =\_\_\_\_\_\_* | | |
| Solution | 0.5645 | | |
| Exercise 13 | *A teacher predicts that the distribution of grades on the final exam will be in Table 11.27.*   |  |  | | --- | --- | | **Grade** | **Proportion** | | A | 0.25 | | B | 0.30 | | C | 0.35 | | D | 0.10 |   *Table 11.27*  *The actual distribution for a class of 20 is in Table 11.28.*   |  |  | | --- | --- | | Grade | Frequency | | A | 7 | | B | 7 | | C | 5 | | D | 1 |   *Table 11.28*  *At the 5% significance level, what can you conclude?* | | |
| Solution | We decline to reject the null hypothesis. There is not enough evidenceto suggest that the observed test scores are significantly different from the expected test scores. | | |
| Exercise 14. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *If the ethnicities of AIDS victims followed the ethnicities of the total county population, fill in the expected number of cases per ethnic group.* | | |
| Solution | |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1748.18 | | **Hispanic** | 26.7% | 1088.03 | | **Black/African-American** | 2.6% | 105.95 | | **Asian, Pacific Islander** | 27.8% | 1132.85 | |  | Total = 100% |  | | | |
| Exercise 15. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 3,464.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *H0:\_\_\_\_\_\_* | | |
| Solution | *H0*: the distribution of AIDS cases follows the ethnicities of the general population of Santa Clara County. | | |
| Exercise 16. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *Ha:\_\_\_\_\_\_* | | |
| Solution | *Ha*: the distribution of AIDS cases does not follow the ethnicities ASI - edited text of the general population of Santa Clara County | | |
| Exercise 17. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.:*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *Is this a right-tailed, left-tailed, or two-tailed test?* | | |
| Solution | right-tailed | | |
| Exercise 18. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *degrees of freedom =* \_\_\_\_\_\_\_ | | |
| Solution | degrees of freedom = 3 | | |
| Exercise 19. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *χ2 test statistic =* \_\_\_\_\_\_\_ | | |
| Solution | 2016.136 | | |
| Exercise 20. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *p*-value *=* \_\_\_\_\_\_\_ | | |
| Solution | 0 | | |
| Exercise 21. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 1,748.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *Graph the situation. Label and scale the horizontal axis. Mark the mean and test statistic. Shade in the region corresponding to the p-value.*    *Figure 11.14*  *Let α = 0.05*  *Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *Reason for the Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *Conclusion (write out in complete sentences): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | | |
| Solution | Graph: Check student’s solution. Decision: Reject the null hypothesis. Reason for the Decision: *p*-value < alpha Conclusion (write out in complete sentences): The make-up of AIDS cases does not fit the ethnicities of the general population of Santa Clara County. | | |
| Exercise 22. | *The following data are real. The cumulative number of AIDS cases reported for Santa Clara County is broken down by ethnicity as in Table 11.29.*   |  |  | | --- | --- | | **Ethnicity** | **Number of Cases** | | White | 2,229 | | Hispanic | 1,157 | | Black/African-American | 457 | | Asian, Pacific Islander | 232 | |  | Total = 4,075 |   *Table 11.29*  *The percentage of each ethnic group in Santa Clara County is as in Table 11.30.*   |  |  |  | | --- | --- | --- | | **Ethnicity** | **Percentage of total county population** | **Number expected (round to two decimal places)** | | **White** | 42.9% | 3,464.18 | | **Hispanic** | 26.7% |  | | **Black/African-American** | 2.6% |  | | **Asian, Pacific Islander** | 27.8% |  | |  | Total = 100% |  |   *Table 11.30*  *Perform a goodness-of-fit test to determine whether the occurrence of AIDS cases follows the ethnicities of the general population of Santa Clara County.*  *Does it appear that the pattern of AIDS cases in Santa Clara County corresponds to the distribution of ethnic groups in this county? Why or why not?* | | |
| Solution | No, the hypothesis test resulted in the conclusion that the make-up of AIDS cases does not follow the ethnicity of the general population in Santa Clara County. | | |
| Exercise 23. | *Determine the appropriate test to be used. A pharmaceutical company is interested in the relationship between age and presentation of symptoms for a common viral infection. A random sample is taken of 500 people with the infection across different age groups.* | | |
| Solution | a test of independence | | |
| Exercise 24. | *Determine the appropriate test to be used. The owner of a baseball team is interested in the relationship between player salaries and team winning percentage. He takes a random sample of 100 players from different organizations.* | | |
| Solution | a test of independence | | |
| Exercise 25. | *Determine the appropriate test to be used. A marathon runner is interested in the relationship between the brand of shoes runners wear and their run times. She takes a random sample of 50 runners and records their run times as well as the brand of shoes they were wearing.* | | |
| Solution | a test of independence | | |
| Exercise 26. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31 shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *State the hypotheses.*  *H0: \_\_\_\_\_\_\_*  *Ha: \_\_\_\_\_\_\_* | | |
| Solution | *H0*: The ticket class chosen is independent of the traveling distance. *Ha*: The ticket class chosen is dependent on the traveling distance. | | |
| Exercise 27. | *Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31 shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *df = \_\_\_\_\_\_* | | |
| Solution | 8 | | |
| Exercise 28. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31**shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *How many passengers are expected to travel between 201 and 300 miles and purchase second-class tickets?* | | |
| Solution | 16.08 | | |
| Exercise 29. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31 shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *How many passengers are expected to travel between 401 and 500 miles and purchase first-class tickets?* | | |
| Solution | 6.6 | | |
| Exercise 30. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31 shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *What is the test statistic?* | | |
| Solution | 15.92 | | |
| Exercise 31. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken. Table 11.31**shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *What is the p-value?* | | |
| Solution | 0.0435 | | |
| Exercise 32. | *Transit Railroads is interested in the relationship between travel distance and ticket class purchased. A random sample of 200 passengers is taken.* ***Table 11.31*** *shows the results. The railroad wants to know if a passenger’s choice in ticket class is independent of the distance they must travel.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Traveling Distance*** | ***Third class*** | ***Second class*** | ***First class*** | ***Total*** | | *1-100 miles* | *21* | *14* | *6* | *41* | | *101-200 miles* | *18* | *16* | *8* | *42* | | *201-300 miles* | *16* | *17* | *15* | *48* | | *301-400 miles* | *12* | *14* | *21* | *47* | | *401-500 miles* | *6* | *6* | *10* | *22* | | *Total* | *73* | *67* | *60* | *200* |   *Table 11.31*  *What can you conclude at the 5% level of significance?* | | |
| Solution | We reject the null hypothesis. There is sufficient evidence to conclude that the selection of ticket class is dependent on travel distance. | | |
| Exercise 33. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *Complete the table.*   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Smoking Level Per Day** | **African American** | **Native Hawaiian** | **Latino** | **Japanese Americans** | **White** | **TOTALS** | | 1-10 |  |  |  |  |  |  | | 11-20 |  |  |  |  |  |  | | 21-30 |  |  |  |  |  |  | | 31+ |  |  |  |  |  |  | | Totals |  |  |  |  |  |  |   *Table 11.32 Smoking Levels by Ethnicity (observed)* | | |
| Solution | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Smoking Level Per Day** | **African American** | **Native Hawaiian** | **Latino** | **Japanese Americans** | **White** | **Totals** | | 1-10 | 9,886 | 2,745 | 12,831 | 8,378 | 7,650 | 41,490 | | 11-20 | 6,514 | 3,062 | 4,932 | 10,680 | 9,877 | 35,065 | | 21-30 | 1,671 | 1,419 | 1,406 | 4,715 | 6,062 | 15,273 | | 31+ | 759 | 788 | 800 | 2,305 | 3,970 | 8,622 | | Totals | 18,830 | 8,014 | 19,969 | 26,078 | 27,559 | 10,0450 |   Table 11.60 | | |
| Exercise 34. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *State the hypotheses.*  *H0: \_\_\_\_\_\_\_*  *Ha: \_\_\_\_\_\_\_* | | |
| Solution | *H0*: Smoking level is independent of ethnic group. *Ha*: Smoking level is dependent on ethnic group. | | |
| Exercise 35. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Smoking Level Per Day** | **African American** | **Native Hawaiian** | **Latino** | **Japanese Americans** | **White** | **Totals** | | 1-10 |  |  |  |  |  |  | | 11-20 |  |  |  |  |  |  | | 21-30 |  |  |  |  |  |  | | 31+ |  |  |  |  |  |  | | Totals |  |  |  |  |  |  |   *Table 11.32*  *Enter expected values in Table 11.32. Round to two decimal places.* | | |
| Solution | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Smoking Level Per Day** | **African American** | **Native Hawaiian** | **Latino** | **Japanese Americans** | **White** | | 1-10 | 7777.57 | 3310.11 | 8248.02 | 10771.29 | 11383.01 | | 11-20 | 6573.16 | 2797.52 | 6970.76 | 9103.29 | 9620.27 | | 21-30 | 2863.02 | 1218.49 | 3036.20 | 3965.05 | 4190.23 | | 31+ | 1616.25 | 687.87 | 1714.01 | 2238.37 | 2365.49 |   Table 11.61 | | |
| Exercise 36. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *df = \_\_\_\_\_\_* | | |
| Solution | 12 | | |
| Exercise 37. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *χ2  test statistic =\_\_\_\_\_\_* | | |
| Solution | 10,301.8 | | |
| Exercise 38. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *p-value =\_\_\_\_\_\_* | | |
| Solution | 0 | | |
| Exercise 39. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *Is this a right-tailed, left-tailed, or two-tailed test? Explain why.* | | |
| Solution | right | | |
| Exercise 40. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *Graph the situation. Label and scale the horizontal axis. Mark the mean and test statistic. Shade in the region corresponding to the p-value.*  *L:\JOBS\XML Jobs\Backend\WNI-CNX-STATISTICS\Phase 2\Converted Arts\Chapter 11\fig-ch11_01_009.jpg*  *Figure 11.15* | | |
| Solution | Check student’s solution. | | |
| Exercise 41. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *Graph the situation. Label and scale the horizontal axis. Mark the mean and test statistic.*  *State the decision and conclusion (in a complete sentence) for the following preconceived levels of α.*  *α = 0.05*  *a. Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *b. Reason for the decision:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *c. Conclusion (write out in a complete sentence):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | | |
| Solution | a. Reject the null hypothesis.  b. *p*-value < alpha  c. There is sufficient evidence to conclude that smoking level is dependent on ethnic group. | | |
| Exercise 42. | *An article in the New England Journal of Medicine, discussed a study on smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans and 7,650 whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese Americans, and 6,062 whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 whites.*  *Graph the situation. Label and scale the horizontal axis. Mark the mean and test statistic.*  *State the decision and conclusion (in a complete sentence) for the following preconceived levels of α.*  *α = 0.01*  *a. Decision:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *b. Reason for the decision:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*  *c. Conclusion (write out in a complete sentence):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* | | |
| Solution | a. Reject the null hypothesis.  b. *p*-value < alpha  c. There is sufficient evidence to conclude that smoking level is dependent on the ethnic group. | | |
| Exercise 43. | *A math teacher wants to see if two of her classes have the same distribution of test scores. What test should she use?* | | |
| Solution | test for homogeneity | | |
| Exercise 44. | *What are the null and alternative hypotheses for Exercise 11.43?* | | |
| Solution | *H0*: The distribution of test scores for the two classes are the same. *Ha*: The distribution of test scores for the two classes are not the same. | | |
| Exercise 45. | *A market researcher wants to see if two different stores have the same distribution of sales throughout the year. What type of test should he use?* | | |
| Solution | test for homogeneity | | |
| Exercise 46. | *A meteorologist wants to know if East and West Australia have the same distribution of storms. What type of test should she use?* | | |
| Solution | test for homogeneity | | |
| Exercise 47. | *What condition must be met to use the test for homogeneity?* | | |
| Solution | All values in the table must be greater than or equal to five. | | |
| Exercise 48. | *Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. The results are shown in Table 11.33.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **20–30** | **30–40** | **40–50** | **50–60** | | Private Practice | 16 | 40 | 38 | 6 | | Hospital | 8 | 44 | 59 | 39 |   *Table 11.33*  *State the null and alternative hypotheses.* | | |
| Solution | *H0*: The distribution of working hours for private practice doctors is the same as the distribution of working hours for hospital doctors. *Ha*: The distribution of working hours for private practice doctors is not the same as the distribution of working hours for hospital doctors. | | |
| Exercise 49. | *Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. The results are shown in Table 11.33.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **20–30** | **30–40** | **40–50** | **50–60** | | Private Practice | 16 | 40 | 38 | 6 | | Hospital | 8 | 44 | 59 | 39 |   *Table 11.33*  *df = \_\_\_\_\_\_\_* | | |
| Solution | 3 | | |
| Exercise 50. | *Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. The results are shown in Table 11.33.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **20–30** | **30–40** | **40–50** | **50–60** | | Private Practice | 16 | 40 | 38 | 6 | | Hospital | 8 | 44 | 59 | 39 |   *Table 11.33*  *What is the test statistic?* | | |
| Solution | 22.50 | | |
| Exercise 51. | *Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. The results are shown in Table 11.33.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **20–30** | **30–40** | **40–50** | **50–60** | | Private Practice | 16 | 40 | 38 | 6 | | Hospital | 8 | 44 | 59 | 39 |   *Table 11.33*  *What is the p-value?* | | |
| Solution | 0.00005 | | |
| Exercise 52. | *Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. The results are shown in Table 11.33.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **20–30** | **30–40** | **40–50** | **50–60** | | Private Practice | 16 | 40 | 38 | 6 | | Hospital | 8 | 44 | 59 | 39 |   *Table 11.33*  *What can you conclude at the 5% significance level?* | | |
| Solution | We reject the null hypothesis. There is enough evidence to show that the distribution of working hours for private practice doctors is not the same as the distribution of working hours for hospital doctors. | | |
| Exercise 53. | *Which test do you use to decide whether an observed distribution is the same as an expected distribution?* | | |
| Solution | a goodness-of-fit test | | |
| Exercise 54. | *What is the null hypothesis for the type of test from Exercise 11.53?* | | |
| Solution | *H0* : The observed distribution fits the expected distribution. | | |
| Exercise 55. | *Which test would you use to decide whether two factors have a relationship?* | | |
| Solution | a test of independence | | |
| Exercise 56. | *Which test would you use to decide if two populations have the same distribution?* | | |
| Solution | a test for homogeneity | | |
| Exercise 57. | *How are tests of independence similar to tests for homogeneity?* | | |
| Solution | Answers will vary. Sample answer: Tests of independence and tests for homogeneity both calculate the test statistic the same way . In addition, all values must be greater than or equal to five. | | |
| Exercise 58. | *How are tests of independence different from tests for homogeneity?* | | |
| Solution | Answers may vary. Sample answer: A test of independence tries to determine if two factors are independent or related. A test for homogeneity tries to determine whether or not the distributions are the same or different for the two populations. | | |
| Exercise 59. | *An archer’s standard deviation for his hits is six (data is measured in distance from the center of the target). An observer claims the standard deviation is less.*  *What type of test should be used?* | | |
| Solution | a test of a single variance | | |
| Exercise 60. | *An archer’s standard deviation for his hits is six (data is measured in distance from the center of the target). An observer claims the standard deviation is less.*  *State the null and alternative hypotheses.* | | |
| Solution | *H0*: σ2 = 36; *Ha*: σ2 < 36 | | |
| Exercise 61. | *An archer’s standard deviation for his hits is six (data is measured in distance from the center of the target). An observer claims the standard deviation is less.*  *Is this a right-tailed, left-tailed, or two-tailed test?* | | |
| Solution | a left-tailed test | | |
| Exercise 62. | *The standard deviation of heights for students in a school is 0.81. A random sample of 50 students is taken, and the standard deviation of heights of the sample is 0.96. A researcher in charge of the study believes the standard deviation of heights for the school is greater than 0.81.*  *What type of test should be used?* | | |
| Solution | a test of a single variance | | |
| Exercise 63. | *The standard deviation of heights for students in a school is 0.81. A random sample of 50 students is taken, and the standard deviation of heights of the sample is 0.96. A researcher in charge of the study believes the standard deviation of heights for the school is greater than 0.81.*  *State the null and alternative hypotheses.* | | |
| Solution | *H0*: σ2 = 0.812*; Ha*: σ2 > 0.812 | | |
| Exercise 64. | *The standard deviation of heights for students in a school is 0.81. A random sample of 50 students is taken, and the standard deviation of heights of the sample is 0.96. A researcher in charge of the study believes the standard deviation of heights for the school is greater than 0.81.*  *df = \_\_\_\_\_\_\_* | | |
| Solution | 49 | | |
| Exercise 65. | *The average waiting time in a doctor’s office varies. The standard deviation of waiting times in a doctor’s office is 3.4 minutes. A random sample of 30 patients in the doctor’s office has a standard deviation of waiting times of 4.1 minutes. One doctor believes the variance of waiting times is greater than originally thought.*  *What type of test should be used?* | | |
| Solution | a test of a single variance | | |
| Exercise 66. | *The average waiting time in a doctor’s office varies. The standard deviation of waiting times in a doctor’s office is 3.4 minutes. A random sample of 30 patients in the doctor’s office has a standard deviation of waiting times of 4.1 minutes. One doctor believes the variance of waiting times is greater than originally thought.*  *What is the test statistic?* | | |
| Solution | 42.17 | | |
| Exercise 67. | *The average waiting time in a doctor’s office varies. The standard deviation of waiting times in a doctor’s office is 3.4 minutes. A random sample of 30 patients in the doctor’s office has a standard deviation of waiting times of 4.1 minutes. One doctor believes the variance of waiting times is greater than originally thought.*  *What is the p-value?* | | |
| Solution | 0.0542 | | |
| Exercise 68. | *The average waiting time in a doctor’s office varies. The standard deviation of waiting times in a doctor’s office is 3.4 minutes. A random sample of 30 patients in the doctor’s office has a standard deviation of waiting times of 4.1 minutes. One doctor believes the variance of waiting times is greater than originally thought.*  *What can you conclude at the 5% significance level?* | | |
| Solution | We decline to reject the null hypothesis. There is not sufficient evidence to conclude that the standard deviation of waiting times is greater than 3.4. | | |
| Exercise 69. | *Decided whether this statement is true or false. As the number of degrees of freedom increases, the graph of the chi-square distribution looks more and more symmetrical.* | | |
| Solution | true | | |
| Exercise 70. | *Decided whether this statement is true or false. The standard deviation of the chi-square distribution is twice the mean.* | | |
| Solution | false | | |
| Exercise 71. | *Decided whether this statement is true or false. The mean and the median of the chi-square distribution are the same if df = 24.* | | |
| Solution | false | | |
| Exercise 72. | *A six-sided die is rolled 120 times. Fill in the expected frequency column. Then, conduct a hypothesis test to determine if the die is fair. The data in Table 11.34 are the result of the 120 rolls.*   |  |  |  | | --- | --- | --- | | **Face Value** | **Frequency** | **Expected Frequency** | | 1 | 15 |  | | 2 | 29 |  | | 3 | 16 |  | | 4 | 15 |  | | 5 | 30 |  | | 6 | 15 |  |   *Table 11.34* | | |
| Solution | |  |  |  | | --- | --- | --- | | **Face Value** | **Frequency** | **Expected Frequency** | | 1 | 15 | 20 | | 2 | 29 | 20 | | 3 | 16 | 20 | | 4 | 15 | 20 | | 5 | 30 | 20 | | 6 | 15 | 20 |   a. *H0*: The die is fair.  b. *Ha*: The die is not fair.  c. degrees of freedom = 5  d. chi-square distribution with *df* = 5  e. chi-square test statistic = 13.6  f. *p*-value = 0.0184  g. Check student’s solution.  h. Decision: reject the null hypothesis, alpha = 0.05, *p*-value < alpha  Conclusion: The die is not fair. | | |
| Exercise 73 | *The marital status distribution of the U.S. male population, ages 15 and older, is as shown in Table 11.35.*   |  |  |  | | --- | --- | --- | | **Marital Status** | **Percent** | **Expected Frequency** | | never married | 31.3 |  | | married | 56.1 |  | | widowed | 2.5 |  | | divorced/separated | 10.1 |  |   *Table 11.35*  *Suppose that a random sample of 400 U.S. young adult males, 18 to 24 years old, yielded the following frequency distribution. We are interested in whether this age group of males fits the distribution of the U.S. adult population. Calculate the frequency one would expect when surveying 400 people. Fill in Table 11.36, rounding to two decimal places.*   |  |  | | --- | --- | | **Marital Status** | **Frequency** | | never married | 140 | | married | 238 | | widowed | 2 | | divorced/separated | 20 |   *Table 11.36* | | |
| Solution | |  |  |  | | --- | --- | --- | | **Marital Status** | **Percent** | **Expected Frequency** | | never married | 31.3 | 125.2 | | married | 56.1 | 224.4 | | widowed | 2.5 | 10 | | divorced/separated | 10.1 | 40.4 |   Table 11.62  a. The data fits the distribution.  b. The data do not fit the distribution.  c. 3  d. chi-square distribution with *df* = 3  e. 19.27  f. 0.0002  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Reject null  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: Data does not fit the distribution. | | |
| Exercise 74. | *The columns in Table 11.37**contain the Race/Ethnicity of U.S. Public Schools for a recent year, the percentages for the Advanced Placement Examinee Population for that class and the Overall Student Population. Suppose the right column contains the result of a survey of 1,000 local students from that year who took an AP Exam.*   |  |  |  |  | | --- | --- | --- | --- | | **Race/Ethnicity** | **AP Examinee Population** | **Overall Student Population** | **Survey Frequency** | | Asian, Asian American, or Pacific Islander | 10.2% | 5.4% | 113 | | Black or African-American | 8.2% | 14.5% | 94 | | Hispanic or Latino | 15.5% | 15.9% | 136 | | American Indian or Alaska Native | 0.6% | 1.2% | 10 | | White | 59.4% | 61.6% | 604 | | Not reported/other | 6.1% | 1.4% | 43 |   *Table 11.37*  Perform a goodness-of-fit test to determine whether the local results follow the distribution of the U. S. Overall Student Population based on ethnicity. | | |
| Solution | a. *H0*: The local results follow the distribution of the U.S. overall student population  b. *Ha*: The local results do not follow the distribution of the U.S. overall student population  c. 5  d. chi-square distribution with *df* = 5  e. chi-square test statistic = 146.4  f. *p*-value = 0  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Reject null hypothesis  iii. Reason for Decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the local results do not follow the distribution of the U. S. overall student population. | | |
| Exercise 75. | *The columns in Table 11.37 contain the Race/Ethnicity of U.S. Public Schools for a recent year, the percentages for the Advanced Placement Examinee Population for that class and the Overall Student Population. Suppose the right column contains the result of a survey of 1,000 local students from that year who took an AP Exam.*   |  |  |  |  | | --- | --- | --- | --- | | **Race/Ethnicity** | **AP Examinee Population** | **Overall Student Population** | **Survey Frequency** | | Asian, Asian American, or Pacific Islander | 10.2% | 5.4% | 113 | | Black or African-American | 8.2% | 14.5% | 94 | | Hispanic or Latino | 15.5% | 15.9% | 136 | | American Indian or Alaska Native | 0.6% | 1.2% | 10 | | White | 59.4% | 61.6% | 604 | | Not reported/other | 6.1% | 1.4% | 43 |   *Table 11.37*  Perform a goodness-of-fit test to determine whether the local results follow the distribution of U.S. AP examinee population, based on ethnicity. | | |
| Solution | a. *H0*: The local results follow the distribution of the U.S. AP examinee population  b. *Ha*: The local results do not follow the distribution of the U.S. AP examinee population  c. *df* = 5  d. chi-square distribution with *df* = 5  e. chi-square test statistic = 13.4  f. *p*-value = 0.0199  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Reject null when *a* = 0.05  iii. Reason for Decision: *p*-value < alpha  iv. Conclusion: Local data do not fit the AP Examinee Distribution.  v. Decision: Do not reject null when *a* = 0.01  vi. Conclusion: There is insufficient evidence to conclude that local data do not follow the distribution of the U.S. AP examinee distribution. | | |
| *Exercise 76.* | *The City of South Lake Tahoe, CA, has an Asian population of 1,419 people, out of a total population of 23,609. Suppose that a survey of 1,419 self-reported Asians in the Manhattan, NY, area yielded the data in Table 11.38. Conduct a goodness-of-fit test to determine if the self-reported sub-groups of Asians in the Manhattan area fit that of the Lake Tahoe area.*   |  |  |  | | --- | --- | --- | | ***Race*** | ***Lake Tahoe Frequency*** | ***Manhattan Frequency*** | | *Asian Indian* | *131* | *174* | | *Chinese* | *118* | *557* | | *Filipino* | *1,045* | *518* | | *Japanese* | *80* | *54* | | *Korean* | *12* | *29* | | *Vietnamese* | *9* | *21* | | *Other* | *24* | *66* |   *Table 11.38* | | |
| Solution | a. *H0*: The sub-groups of Asians in the Manhattan area fit that of the Lake Tahoe area  b. *Ha*: The sub-groups of Asians in the Manhattan area do not fit that of the Lake Tahoe area  c. *df* = 6  d. chi-square distribution with *df* = 6  e. test statistic = 2035  f. *p*-value =0  g. Check student’s solution.  h.  i. Alpha = 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the sub-groups of Asians in the Manhattan area do not fit that of the Lake Tahoe area. | | |
| Exercise 77. | *UCLA conducted a survey of more than 263,000 college freshmen from 385 colleges in fall 2005. The results of students' expected majors by gender were reported in The Chronicle of Higher Education (2/2/2006). Suppose a survey of 5,000 graduating females and 5,000 graduating males was done as a follow-up last year to determine what their actual majors were. The results are shown in the tables for Exercise 11.77 and Exercise 11.78. The second column in each table does not add to 100% because of rounding.*  *Conduct a goodness-of-fit test to determine if the actual college major of graduating females fits the distribution of their expected majors.*   |  |  |  | | --- | --- | --- | | **Major** | **Women - Expected Major** | **Women – Actual Major** | | Arts & Humanities | 14.0% | 670 | | Biological Sciences | 8.4% | 410 | | Business | 13.1% | 685 | | Education | 13.0% | 650 | | Engineering | 2.6% | 145 | | Physical Sciences | 2.6% | 125 | | Professional | 18.9% | 975 | | Social Sciences | 13.0% | 605 | | Technical | 0.4% | 15 | | Other | 5.8% | 300 | | Undecided | 8.0% | 420 |   *Table 11.39* | | |
| Solution | a. *H0*: The actual college majors of graduating females fit the distribution of their expected majors  b. *Ha*: The actual college majors of graduating females do not fit the distribution of their expected majors  c. *df* = 10  d. chi-square distribution with *df* = 10  e. test statistic = 11.48  f. *p*-value = 0.3211  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Do not reject null when *a* = 0.05 and *a* = 0.01  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the distribution of actual college majors of graduating females fits the distribution of their expected majors. | | |
| Exercise 78. | *UCLA conducted a survey of more than 263,000 college freshmen from 385 colleges in fall 2005. The results of students' expected majors by gender were reported in The Chronicle of Higher Education (2/2/2006). Suppose a survey of 5,000 graduating females and 5,000 graduating males was done as a follow-up last year to determine what their actual majors were. The results are shown in the tables for Exercise 11.77 and Exercise 11.78. The second column in each table does not add to 100% because of rounding.*  *Conduct a goodness-of-fit test to determine if the actual college majors of graduating males fits the distribution of their expected majors.*   |  |  |  | | --- | --- | --- | | **Major** | **Men - Expected Major** | **Men – Actual Major** | | Arts & Humanities | 11.0% | 600 | | Biological Sciences | 6.7% | 330 | | Business | 22.7% | 1130 | | Education | 5.8% | 305 | | Engineering | 15.6% | 800 | | Physical Sciences | 3.6% | 175 | | Professional | 9.3% | 460 | | Social Sciences | 7.6% | 370 | | Technical | 1.8% | 90 | | Other | 8.2% | 400 | | Undecided | 6.6% | 340 |   *Table 11.40* | | |
| Solution | a. *Ho*: The actual college majors of graduating males fit the distribution of their expected majors  b. *Ha*: The actual college majors of graduating males do not fit the distribution of their expected majors  c. *df* = 10  d. chi-square distribution with *df* = 10  e. test statistic = 6.93  f. *p*-value = 0.7317  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the distribution of actual college majors of graduating males fits the distribution of their expected majors. | | |
| Exercise 79. | *Read the statement and decide whether it is true or false.*  *In a goodness-of-fit test, the expected values are the values we would expect if the null hypothesis were true.* | | |
| Solution | true | | |
| Exercise 80. | *Read the statement and decide whether it is true or false.*  *In general, if the observed values and expected values of a goodness-of-fit test are not close together, then the test statistic can get very large and on a graph will be way out in the right tail.* | | |
| Solution | true | | |
| Exercise 81. | *Read the statement and decide whether it is true or false.*  *Use a goodness-of-fit test to determine if high school principals believe that students are absent equally during the week or not.* | | |
| Solution | true | | |
| Exercise 82. | *Read the statement and decide whether it is true or false.*  *The test to use to determine if a six-sided die is fair is a goodness-of-fit test.* | | |
| Solution | true | | |
| Exercise 83. | *Read the statement and decide whether it is true or false.*  *In a goodness-of fit test, if the p-value is 0.0113, in general, do not reject the null hypothesis.* | | |
| Solution | false | | |
| Exercise 84. | *A sample of 212 commercial businesses was surveyed for recycling one commodity; a commodity here means any one type of recyclable material such as plastic or aluminum. Table 11.41 shows the business categories in the survey, the sample size of each category, and the number of businesses in each category that recycle one commodity. Based on the study, on average half of the businesses were expected to be recycling one commodity. As a result, the last column shows the expected number of businesses in each category that recycle one commodity. At the 5% significance level, perform a hypothesis test to determine if the observed number of businesses that recycle one commodity follows the uniform distribution of the expected values.*   |  |  |  |  | | --- | --- | --- | --- | | **Business Type** | **Number in class** | **Observed Number that recycle one commodity** | **Expected number that recycle one commodity** | | Office | 35 | 19 | 17.5 | | Retail/Wholesale | 48 | 27 | 24 | | Food/Restaurants | 53 | 35 | 26.5 | | Manufacturing/Medical | 52 | 21 | 26 | | Hotel/Mixed | 24 | 9 | 12 |   *Table 11.41* | | |
| Solution | *a. H0*: The distribution of businesses that recycle one commodity fits the expected number that recycle one commodity.  *b. Ha*: The distribution of businesses that recycle one commodity does not fit the expected number that recycle one commodity.  c. *df* = 4  d. chi-square distribution with *df* = 4  e. test statistic = 4.9415  f. *p*-value = 0.2934  g.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: At the 5% level of significance, from the data, there is insufficient evidence to conclude that the distribution of businesses that recycle one commodity fit the expected number that recycle one commodity. | | |
| Exercise 85. | *Table 11.42**contains information from a survey among 499 participants classified according to their age groups. The second column shows the percentage of obese people per age class among the study participants. The last column comes from a different study at the national level that shows the corresponding percentages of obese people in the same age classes in the USA. Perform a hypothesis test at the 5% significance level to determine whether the survey participants are a representative sample of the USA obese population.*   |  |  |  | | --- | --- | --- | | **Age Class (Years)** | **Obese (Percentage)** | **Expected USA average**  **(Percentage)** | | 20–30 | 75.0 | 32.6 | | 31–40 | 26.5 | 32.6 | | 41–50 | 13.6 | 36.6 | | 51–60 | 21.9 | 36.6 | | 61–70 | 21.0 | 39.7 |   *Table 11.42* | | |
| Solution | a. *H0*: Surveyed obese fit the distribution of expected obese  *b. Ha*: Surveyed obese do not fit the distribution of expected obese  c. *df* = 4  d. chi-square distribution with *df* = 4  e. test statistic = 54.01  f. *p*-value = 0  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: At the 5% level of significance, from the data, there is sufficient evidence to conclude that the surveyed obese do not fit the distribution of expected obese. | | |
| Exercise 86. | *A recent debate about where in the United States skiers believe the skiing is best prompted the following survey. Test to see if the best ski area is independent of the level of the skier.*   |  |  |  |  | | --- | --- | --- | --- | | **U.S. Ski Area** | **Beginner** | **Intermediate** | **Advanced** | | Tahoe | 20 | 30 | 40 | | Utah | 10 | 30 | 60 | | Colorado | 10 | 40 | 50 |   *Table 11.43* | | |
| Solution | | a. *H0:* Ski area is independent of the level of the skier.  b. *Ha:* Ski area is dependent on the level of the skier.  c. *df* = 4  d. chi-square distribution with *df* = 4  e. test statistic = 10.53  f. *p*-value = 0.0324  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject null.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: Best ski area and level of skier are not independent. | |
| Exercise 87. | | *Car manufacturers are interested in whether there is a relationship between the size of car an individual drives and the number of people in the driver’s family (that is, whether car size and family size are independent). To test this, suppose that 800 car owners were randomly surveyed with the results in* ***Table 1.37****. Conduct a test of independence.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Family Size** | **Sub & Compact** | **Mid-size** | **Full-size** | **Van & Truck** | | 1 | 20 | 35 | 40 | 35 | | 2 | 20 | 50 | 70 | 80 | | 3–4 | 20 | 50 | 100 | 90 | | 5+ | 20 | 30 | 70 | 70 |   *Table 11.44* | |
| Solution | | a. *H0*: Car size is independent of family size.  b. *Ha*: Car size is dependent on family size.  c. *df* = 9  d. chi-square distribution with *df* = 9  e. test statistic = 15.8284  f. *p*-value = 0.0706  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that car size and family size are dependent. | |
| Exercise 88. | | *College students may be interested in whether or not their majors have any effect on starting salaries after graduation. Suppose that 300 recent graduates were surveyed as to their majors in college and their starting salaries after graduation. Table 11.45 shows the data. Conduct a test of independence.*   |  |  |  |  | | --- | --- | --- | --- | | **Major** | **< $50,000** | **$50,000–$68,999** | **$69,000+** | | English | 5 | 20 | 5 | | Engineering | 10 | 30 | 60 | | Nursing | 10 | 15 | 15 | | Business | 10 | 20 | 30 | | Psychology | 20 | 30 | 20 |   *Table 11.45* | |
| Solution | | a. *H0*: Salaries are independent of majors.  b. *Ha*: Salaries are dependent on majors.  c. *df* = 8  d. chi-square distribution with *df* = 8  e. test statistic = 33.55  f. *p*-value = 0  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Reject the null. 3. Reason for decision: *p*-value < alpha 4. Conclusion: Major and starting salary are not independent events. | |
| Exercise 89. | | *Some travel agents claim that honeymoon hot spots vary according to age of the bride. Suppose that 280 recent brides were interviewed as to where they spent their honeymoons. The information is given in Table 11.46. Conduct a test of independence.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Location** | **20–29** | **30–39** | **40–49** | **50 and over** | | Niagara Falls | 15 | 25 | 25 | 20 | | Poconos | 15 | 25 | 25 | 10 | | Europe | 10 | 25 | 15 | 5 | | Virgin Islands | 20 | 25 | 15 | 5 |   *Table 11.46* | |
| Solution | | a. *H0*: Honeymoon locations are independent of bride’s age.  b. *Ha*: Honeymoon locations are dependent on bride’s age.  c. *df* = 9  d. chi-square distribution with *df* = 9  e. test statistic = 15.7027  f. *p*-value = 0.0734  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Do not reject the null hypothesis. 3. Reason for decision: *p*-value > alpha 4. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that honeymoon location and bride age are dependent. | |
| Exercise 90 | | *A manager of a sports club keeps information concerning the main sport in which members participate and their ages. To test whether there is a relationship between the age of a member and his or her choice of sport, 643 members of the sports club are randomly selected. Conduct a test of independence.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Sport** | **18–25** | **26–30** | **31–40** | **41 and over** | | racquetball | 42 | 58 | 30 | 46 | | tennis | 58 | 76 | 38 | 65 | | swimming | 72 | 60 | 65 | 33 |   *Table 11.47* | |
| Solution | | a. *H0*: The age of a member and his or her main sport are independent.  b. *Ha*: The age of a member and his or her main sport are dependent.  c. *df* = 6  d. chi-square distribution with *df* = 6  e. test statistic = 25.21  f. *p*-value = 0.0003  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Reject the null hypothesis. 3. Reason for decision: *p*-value < alpha 4. Conclusion: At the 5% significance level, there is sufficient evidence to conclude that sport and age are dependent. | |
| Exercise 91. | | *A major food manufacturer is concerned that the sales for its skinny french fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are shown in Table 11.48. Conduct a test of independence.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Type of Fries** | **Northeast** | **South** | **Central** | **West** | | skinny fries | 70 | 50 | 20 | 25 | | curly fries | 100 | 60 | 15 | 30 | | steak fries | 20 | 40 | 10 | 10 |   *Table 11.48* | |
| Solution | | a. *H*0: The types of fries sold are independent of the location.  b. Ha: The types of fries sold are dependent on the location.  c. *df* = 6  d. chi-square distribution with *df* = 6  e. test statistic =18.8369  f. *p*-value = 0.0044  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Reject the null hypothesis. 3. Reason for decision: *p*-value < alpha 4. Conclusion: At the 5% significance level, There is sufficient evidence that types of fries and location are dependent. | |
| Exercise 92. | | *According to Dan Lenard, an independent insurance agent in the Buffalo, N.Y. area, the following is a breakdown of the amount of life insurance purchased by males in the following age groups. He is interested in whether the age of the male and the amount of life insurance purchased are independent events. Conduct a test for independence.*   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Age of Males** | **None** | **<$200,000** | **$200,000–$400,000** | **$400,001–$1,000,000** | **$1,000,000+** | | 20–29 | 40 | 15 | 40 | 0 | 5 | | 30–39 | 35 | 5 | 20 | 20 | 10 | | 40–49 | 20 | 0 | 30 | 0 | 30 | | 50+ | 40 | 30 | 15 | 15 | 10 |   *Table 11.49* | |
| Solution | | a. *H0*: The amount of life insurance is independent of age.  b. *Ha*: The amount of life insurance is dependent on age.  c. *df* = 12  d. chi--square distribution with *df* = 12  e. test statistic = 1 25.7 4  f. *p*-value = 0  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Reject null 3. Reason for decision: *p*-value < alpha 4. Conclusion: At the 5% significance level, there is sufficient evidence to conclude that amount of life insurance and age are dependent. | |
| Exercise 93 | | *Suppose that 600 thirty-year-olds were surveyed to determine whether or not there is a relationship between the level of education an individual has and salary. Conduct a test of independence.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Annual Salary** | **Not a high school graduate** | **High school graduate** | **College graduate** | **Master or doctorate** | | < $30,000 | 15 | 25 | 10 | 5 | | $30,000–$40,000 | 20 | 40 | 70 | 30 | | $40,000–$50,000 | 10 | 20 | 40 | 55 | | $50,000–$60,000 | 5 | 10 | 20 | 60 | | $60,000+ | 0 | 5 | 10 | 150 |   *Table 11.50* | |
| Solution | | a. *H0*: Salary is independent of level of education.  b. *Ha*: Salary is dependent on level of education.  c. *df* = 12  d. chi-square distribution with *df* = 12  e. test statistic = 255.7704  f. *p*-value = 0  g. Check student’s solution.  h.   1. Alpha: 0.05 2. Decision: Reject the null hypothesis. 3. Reason for decision: *p*-value < alpha 4. Conclusion: At the 5% significance level, there is sufficient evidence to conclude that salary and level of education are dependent. | |
| Exercise 94. | | *Read the statement and decide whether it is true or false.*  *The number of degrees of freedom for a test of independence is equal to the sample size minus one.* | |
| Solution | | false | |
| Exercise 95. | | *Read the statement and decide whether it is true or false.*  *The test for independence uses tables of observed and expected data values.* | |
| Solution | | true | |
| Exercise 96. | | *Read the statement and decide whether it is true or false.*  *The test to use when determining if the college or university a student chooses to attend is related to his or her socioeconomic status is a test for independence.* | |
| Solution | | true | |
| Exercise 97. | | *Read the statement and decide whether it is true or false.*  *In a test of independence, the expected number is equal to the row total multiplied by the column total divided by the total surveyed.* | |
| Solution | | true | |
| Exercise 98. | | *An ice cream maker performs a nationwide survey about favorite flavors of ice cream in different geographic areas of the U.S. Based on Table 11.51, do the numbers suggest that geographic location is independent of favorite ice cream flavors? Test at the 5% significance level.*   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **U.S. region/ Flavor** | **Strawberry** | **Chocolate** | **Vanilla** | **Rocky Road** | **Mint Chocolate Chip** | **Pistachio** | **Row Total** | | West | 12 | 21 | 22 | 19 | 15 | 8 | 97 | | Midwest | 10 | 32 | 22 | 11 | 15 | 6 | 96 | | East | 8 | 31 | 27 | 8 | 15 | 7 | 96 | | South | 15 | 28 | 30 | 8 | 15 | 6 | 102 | | Column Total | 45 | 112 | 101 | 46 | 60 | 27 | 391 |   *Table 11.51* | | |
| Solution | | a. *H0*: Favorite Ice Cream Flavor is independent of Geographic Location  b. *Ha*: Favorite Ice Cream Flavor is dependent on Geographic Location  c. *df* = 15  d. chi-square distribution with *df* = 15  e. test statistic = 14.06  f. *p*-value = 0.5207  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for the decision: p-value > α  iv. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that favorite ice cream flavor is dependent on geographic location. | |
| Exercise 99. | | *Table 11.52 provides a recent survey of the youngest online entrepreneurs whose net worth is estimated at one million dollars or more. Their ages range from 17 to 30. Each cell in the table illustrates the number of entrepreneurs who correspond to the specific age group and their net worth. Are the ages and net worth independent? Perform a test of independence at the 5% significance level.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Age Group\Net Worth Value (in millions of US dollars)** | **1**–**5** | **6**–**24** | **≥25** | **Row Total** | | 17-25 | 8 | 7 | 5 | 20 | | 26-30 | 6 | 5 | 9 | 20 | | Column Total | 14 | 12 | 14 | 40 |   *Table 11.52* | |
| Solution | | a. *H0*: Age is independent of the youngest online entrepreneurs’ net worth.  b. *Ha*: Age is dependent on the net worth of the youngest online entrepreneurs.  c. *df* = 2  d. chi-square distribution with *df* = 2  e. test statistic = 1.76  f. *p*-value 0.4144  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that age and net worth for the youngest online entrepreneurs are dependent. | |
| Exercise 100. | | *A 2013 poll in California surveyed people about taxing sugar-sweetened beverages. The results are presented in Table 11.53, and are classified by ethnic group and response type. Are the poll responses independent of the participants’ ethnic group? Conduct a test of independence at the 5% significance level.*   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Opinion/Ethnicity** | **Asian American** | **White/ Non-Hispanic** | **African American** | **Latino** | **Row Total** | | Against tax | 48 | 433 | 41 | 160 | 682 | | In Favor of tax | 54 | 234 | 24 | 147 | 459 | | No opinion | 16 | 43 | 16 | 19 | 94 | | Column Total | 118 | 710 | 81 | 326 | 1235 |   *Table 11.53* | |
| Solution | | a. *H0*: Ethnic group is independent of the response type  *b. Ha*: Ethnic group is dependent on the response type  c. *df* = 6  d. chi-square distribution with *df* = 6  e. test statistic = 54.27  f. *p*-value = 0  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: At the 5% level of significance, there is sufficient evidence to conclude that ethnic group and response type are dependent. | |
| Exercise 101. | | *A Psychologist is interested in testing whether there is a difference in the distribution of personality types for business majors and social science majors. The results of the study are shown in Table 11.54. Conduct a test of homogeneity. Test at a 5% level of significance.*   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | **Open** | **Conscientious** | **Extrovert** | **Agreeable** | **Neurotic** | | **Business** | 41 | 52 | 46 | 61 | 58 | | **Social Science** | 72 | 75 | 63 | 80 | 65 |   *Table 11.54* | |
| Solution | | a. *H0*: The distribution for personality types is the same for both majors  b. *Ha*: The distribution for personality types is not the same for both majors  c. *df* = 4  d. chi-square with *df* = 4  e. test statistic = 3.01  f. *p*-value = 0.5568  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the distribution of personality types is different for business and social science majors. | |
| Exercise 102. | | *Do men and women select different breakfasts? The breakfasts ordered by randomly selected men and women at a popular breakfast place is shown in Table 11.55. Conduct a test for homogeneity at a 5% level of significance.*   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | **French Toast** | **Pancakes** | **Waffles** | **Omelettes** | | **Men** | 47 | 35 | 28 | 53 | | **Women** | 65 | 59 | 55 | 60 |   *Table 11.55* | |
| Solution | | a. *H0*: The distribution for breakfast preferences is the same for men and women  b. *Ha*: The distribution for breakfast preferences is not the same for men and women  c. 3  d. chi-square with *df* = 3  e. 4.01  f. p-value = 0.2601  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the distribution of breakfast ordered is different for men and women. | |
| Exercise 103. | | *A fisherman is interested in whether the distribution of fish caught in Green Valley Lake is the same as the distribution of fish caught in Echo Lake. Of the 191 randomly selected fish caught in Green Valley Lake, 105 were rainbow trout, 27 were other trout, 35 were bass, and 24 were catfish. Of the 293 randomly selected fish caught in Echo Lake, 115 were rainbow trout, 58 were other trout, 67 were bass, and 53 were catfish. Perform a test for homogeneity at a 5% level of significance.* | |
| Solution | | a. *H0*: The distribution for fish caught is the same in Green Valley Lake and in Echo Lake.  b. *Ha*: The distribution for fish caught is not the same in Green Valley Lake and in Echo Lake.  c. 3  d. chi-square with *df* = 3  e. 11.75  f. *p*-value = 0.0083  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: There is evidence to conclude that the distribution of fish caught is different in Green Valley Lake and in Echo Lake | |
| Exercise 104. | | *In 2007, the United States had 1.5 million homeschooled students, according to the U.S. National Center for Education Statistics. In Table 11.56 you can see that parents decide to homeschool their children for different reasons, and some reasons are ranked by parents as more important than others. According to the survey results shown in the table, is the distribution of applicable reasons the same as the distribution of the most important reason? Provide your assessment at the 5% significance level. Did you expect the result you obtained?*   |  |  |  |  | | --- | --- | --- | --- | | **Reasons for Homeschooling** | **Applicable Reason (in thousands of respondents)** | **Most Important Reason (in thousands of respondents)** | **Row Total** | | Concern about the environment of other schools | 1,321 | 309 | 1,630 | | Dissatisfaction with academic instruction at other schools | 1,096 | 258 | 1,354 | | To provide religious or moral instruction | 1,257 | 540 | 1,797 | | Child has special needs, other than physical or mental | 315 | 55 | 370 | | Nontraditional approach to child’s education | 984 | 99 | 1,083 | | Other reasons (e.g., finances, travel, family time, etc.) | 485 | 216 | 701 | | Column Total | 5,458 | 1,477 | 6,935 |   *Table 11.56* | |
| Solution | | a. *H0*: The distribution of applicable reasons is the same as that of the most important reason.  b. *Ha*: The distribution of applicable reasons is not the same as that of the most important reason.  c*. df* = 5  d. chi-square with *df* = 5  e. test statistic = 234  f. *p*-value = 0  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for Decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the Applicable and Most Important reasons for homeschooling come from different distributions. | |
| Exercise 105. | | *When looking at energy consumption, we are often interested in detecting trends over time and how they correlate among different countries. The information in Table 11.57 shows the average energy use (in units of kg of oil equivalent per capita) in the USA and the joint European Union countries (EU) for the six-year period 2005 to 2010. Do the energy use values in these two areas come from the same distribution? Perform the analysis at the 5% significance level.*   |  |  |  |  | | --- | --- | --- | --- | | **Year** | **European Union** | **United States** | **Row Total** | | 2010 | 3,413 | 7,164 | 10,557 | | 2009 | 3,302 | 7,057 | 10,359 | | 2008 | 3,505 | 7,488 | 10,993 | | 2007 | 3,537 | 7,758 | 11,295 | | 2006 | 3,595 | 7,697 | 11,292 | | 2005 | 3,613 | 7,847 | 11,460 | | Column Total | 20,965 | 45,011 | 65,976 |   *Table 11.57* | |
| Solution | | a. *H0*: The distribution of average energy use in the USA is the same as in Europe between 2005 and 2010.  b. *Ha*: The distribution of average energy use in the USA is not the same as in Europe between 2005 and 2010.  c. *df* = 4  d. chi-square with *df* = 4  e. test statistic = 2.7434  f. *p*-value = 0.7395  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that the average energy use values in the US and EU are not derived from different distributions for the period from 2005 to 2010. | |
| Exercise 106. | | *The Insurance Institute for Highway Safety collects safety information about all types of cars every year, and publishes a report of Top Safety Picks among all cars, makes, and models. Table 11.58 presents the number of Top Safety Picks in six car categories for the two years 2009 and 2013. Analyze the table data to conclude whether the distribution of cars that earned the Top Safety Picks .*   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Year \ Car Type** | **Small** | **Mid-Size** | **Large** | **Small**  **SUV** | **Mid-Size**  **SUV** | **Large**  **SUV** | **Row**  **Total** | | 2009 | 12 | 22 | 10 | 10 | 27 | 6 | 87 | | 2013 | 31 | 30 | 19 | 11 | 29 | 4 | 124 | | **Column Total** | 43 | 52 | 29 | 21 | 56 | 10 | 211 |   *Table 11.58* | |
| Solution | | a*. H0*: The distribution of Top Safety Picks cars in 2009 is the same as the corresponding 2013 distribution.  b*. Ha*: The distribution of Top Safety Picks cars in 2009 is not the same as the corresponding 2013 distribution.  c. *df* = 5  d. chi-square with *df* = 5  e. test statistic = 6.6547  f. *p*-value = 0.2476  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject he null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: At the 5% significance level, there is insufficient evidence to conclude that the two distributions are different for the years 2009 and 2013.  Under a different light, with the exceptions of the large SUV category, all car categories were awarded an increased number of safer cars in 2013. The present analysis suggest that car companies have taken care of safety in their newly awarded models proportionally across categories, so that the 2013 distribution of Top Safety Picks has remained the same as the corresponding 2009 distribution. | |
| Exercise 107. | | *Is there a difference between the distribution of community college statistics students and the distribution of university statistics students in what technology they use on their homework? Of some randomly selected community college students, 43 used a computer, 102 used a calculator with built in statistics functions, and 65 used a table from the textbook. Of some randomly selected university students, 28 used a computer, 33 used a calculator with built in statistics functions, and 40 used a table from the textbook. Conduct an appropriate hypothesis test using a 0.05 level of significance.* | |
| Solution | | a. *H0*: The distribution for technology use is the same for community college students and university students.  b. *Ha*: The distribution for technology use is not the same for community college students and university students.  c. 2  d. chi-square with *df* =2  e. 7.05  f. *p*-value = 0.0294  g. Check student’s solution.  h: i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv . Conclusion: There is sufficient evidence to conclude that the distribution of technology use for statistics homework is not the same for statistics students at community colleges and at universities. | |
| Exercise 108. | | *Read the statement and decide whether it is true or false.*  *If df = 2, the chi-square distribution has a shape that reminds us of the exponential.* | |
| Solution | | true | |
| Exercise 109. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *Is the traveler disputing the claim about the average or about the variance?* | |
| Solution | | | Variance | |
| Exercise 110. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *A sample standard deviation of 15 minutes is the same as a sample variance of \_\_\_\_\_\_\_\_\_\_ minutes.* | |
| Solution | | | 225 | |
| Exercise 111. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *Is this a right-tailed, left-tailed, or two-tailed test?* | |
| Solution | | | right-tailed | |
| Exercise 112. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *H0: \_\_\_\_\_\_\_\_\_\_* | |
| Solution | | | *H0*: σ2 ≤ 150 | |
| Exercise 113. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *df = \_\_\_\_\_\_\_\_* | |
| Solution | | | 24 | |
| Exercise 114. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *chi-square test statistic = \_\_\_\_\_\_\_\_* | |
| Solution | | | 36 | |
| Exercise 115. | | | *Suppose an airline claims that its flights are consistently on time with an average delay of at most 15 minutes. It claims that the average delay is so consistent that the variance is no more than 150 minutes. Doubting the consistency part of the claim, a disgruntled traveler calculates the delays for his next 25 flights. The average delay for those 25 flights is 22 minutes with a standard deviation of 15 minutes.*  *p-value = \_\_\_\_\_\_\_\_* | |
| Solution | | | 0.0549 | |
| Exercise 116. | | | *Graph the situation. Label and scale the horizontal axis. Mark the mean and test statistic. Shade the p-value.* | |
| Solution | | | Check student’s solution. | |
| Exercise 117. | | | *Let α = 0.05*  *Decision:\_\_\_\_\_\_\_\_\_\_*  *Conclusion (write out in a complete sentence.):\_\_\_\_\_\_\_\_\_\_* | |
| Solution | | | Decision: Do not reject the null hypothesis.  Conclusion: There is not sufficient evidence to conclude that the variation of flight delay times is more than 150 minutes. | |
| Exercise 118. | | | *How did you know to test the variance instead of the mean?* | |
| Solution | | | The claim is that the variance is no more than 150 minutes. | |
| Exercise 119. | | | *If an additional test were done on the claim of the average delay, which distribution would you use?* | |
| Solution | | | a Student’s  *t*-distribution | |
| Exercise 120. | | | *If an additional test were done on the claim of the average delay, which distribution would you use?* | |
| Solution | | | a Student’s *t*- or normal distribution | |
| Exercise 121. | | | *A plant manager is concerned her equipment may need recalibrating. It seems that the actual weight of the 15 oz. cereal boxes it fills has been fluctuating. The standard deviation should be at most 0.5 oz. In order to determine if the machine needs to be recalibrated, 84 randomly selected boxes of cereal from the next day’s production were weighed. The standard deviation of the 84 boxes was 0.54. Does the machine need to be recalibrated?* |
| Solution | | | a. H0: σ ≤ 0.5  b. Ha: σ > 0.5  c. *df* = 83  d. chi-square with *df* = 83  e. test statistic = 96.81  f. *p*-value = 0.1426; There is a 0.1426 probability that the sample standard deviation is 0.54 or more.  g. Check student’s solution.  h. i. Alpha = 0.05.  ii. Decision: Do not reject null  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the standard deviation is more than 0.5 oz. It cannot be determined whether the equipment needs to be recalibrated or not. |
| Exercise 122. | | | *Consumers may be interested in whether the cost of a particular calculator varies from store to store. Based on surveying 43 stores, which yielded a sample mean of $84 and a sample standard deviation of $12, test the claim that the standard deviation is greater than $15.* |
| Solution | | | a. *H0:* σ = 15  b. *Ha:* σ > 15  c. *df* = 42  d. chi-square with *df* = 42  e. test statistic =26.88  f. *p*-value = 0.9663  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Do not reject null hypothesis.  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the standard deviation is greater than 15. |
| Exercise 123. | | | *Isabella, an accomplished Bay to Breakers**runner, claims that the standard deviation for her time to run the 7.5 mile race is at most three minutes. To test her claim, Rupinder looks up five of her race times. They are 55 minutes, 61 minutes, 58 minutes, 63 minutes, and 57 minutes.* |
| Solution | | | a. *H0:* σ ≤ 3  b. *Ha:* σ > 3  c. *df* = 4  d. chi-square with *df* = 4  e. test statistic = 4.52  f. *p*-value =0.3402  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Do not reject null.  iii. Reason for decision: : *p*-value > alpha  iv. Conclusion: There is not sufficient evidence to conclude that Isabella’s standard deviation for her time to run the Bay to Breaker’s race is greater than 3 minutes. |
| Exercise 124. | | | *Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. They are also interested in the variation of the number of babies. Suppose that an airline executive believes the average number of babies on flights is six with a variance of nine at most. The airline conducts a survey. The results of the 18 flights surveyed give a sample average of 6.4 with a sample standard deviation of 3.9. Conduct a hypothesis test of the airline executive’s belief.* |
| Solution | | | a. *H0:* σ ≤ 3  b. *Ha:* σ > 3  c. *df* = 17  d. chi-square distribution with *df* = 17  e. test statistic = 28.73  f. *p*-value = 0.0371  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the standard deviation is greater than three. |
| Exercise 125. | | | *The number of births per woman in China is 1.6 down from 5.91 in 1966. This fertility rate has been attributed to the law passed in 1979 restricting births to one per woman. Suppose that a group of students studied whether or not the standard deviation of births per woman was greater than 0.75. They asked 50 women across China the number of births they had had. The results are shown in Table 11.59. Does the students’ survey indicate that the standard deviation is greater than 0.75?*   |  |  | | --- | --- | | ***# of births*** | ***Frequency*** | | *0* | *5* | | *1* | *30* | | *2* | *10* | | *3* | *5* |   *Table 11.59* |
| Solution | | | a. *H0*: σ ≤ 0.75  b. *Ha*: σ > 0.75  c. *df* = 49  d. chi-square with *df* = 49  e. test statistic = 54.37  f. *p*-value = 0.2774; If the null hypothesis is true, there is a 0.2774 probability that the sample standard deviation is 0.79 or more.  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Do not reject the null  iii. Reason for decision: *p*-value > alpha  iv. Conclusion: There is insufficient evidence to conclude that the standard deviation is greater than 0.75. |
| Exercise 126. | | | *According to an avid aquarist, the average number of fish in a 20-gallon tank is ten, with a standard deviation of two. His friend, also an aquarist, does not believe that the standard deviation is two. She counts the number of fish in 15 other 20-gallon tanks. Based on the results that follow, do you think that the standard deviation is different from two? Data: 11; 10; 9; 10; 10; 11; 11; 10; 12; 9; 7; 9; 11; 10; 11* |
| Solution | | | a. *H0*: σ = 2  b. *Ha*: σ ≠ 2  c. *df* = 14  d. chi-Square distribution with *df* = 14  e. chi-square test statistic = 5.2094  f. *p*-value = 0.0346  g. Check student’s solution.  h. i. Alpha = 0.05  ii. Decision: Reject the null hypothesis.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the standard deviation is different than 2. |
| Exercise 127. | | | *The manager of "Frenchies" is concerned that patrons are not consistently receiving the same amount of French fries with each order. The chef claims that the standard deviation for a ten-ounce order of fries is at most 1.5 oz., but the manager thinks that it may be higher. He randomly weighs 49 orders of fries, which yields a mean of 11 oz. and a standard deviation of two oz.* |
| Solution | | | a*. H0*: σ2 ≤ 1.5  b. *Ha*: σ > 1.5  c. 48  d. chi-square with *df* = 48  e. 85.33  f. 0.0007  g. Check student’s solution.  h. i. Alpha: 0.05  ii. Decision: Reject null.  iii. Reason for decision: *p*-value < alpha  iv. Conclusion: There is sufficient evidence to conclude that the standard deviation for a ten-ounce order of fries is more than 1.5 ounces. |
| Exercise 128. | | | *You want to buy a specific computer. A sales representative of the manufacturer claims that retail stores sell this computer at an average price of $1,249 with a very narrow standard deviation of $25. You find a website that has a price comparison for the same computer at a series of stores as follows: $1,299; $1,229.99; $1,193.08; $1,279; $1,224.95; $1,229.99; $1,269.95; $1,249. Can you argue that pricing has a larger standard deviation than claimed by the manufacturer? Use the 5% significance level. As a potential buyer, what would be the practical conclusion from your analysis?* | |
| Solution | | | The sample standard deviation is $34.29. *H0*: *σ 2* = 252  *Ha*: *σ 2* > 252  *df* = *n* – 1 = 7.  test statistic: ;  *p*-value:  Alpha: 0.05  Decision: Do not reject the null hypothesis.  Reason for decision: p-value > alpha  Conclusion: At the 5% level, there is insufficient evidence to conclude that the variance is more than 625. | |
| Exercise 129. | | | *A company packages apples by weight. One of the weight grades is Class A apples. Class A apples have a mean weight of 150 g, and there is a maximum allowed weight tolerance of 5% above or below the mean for apples in the same consumer package. A batch of apples is selected to be included in a Class A apple package. Given the following apple weights of the batch, does the fruit comply with the Class A grade weight tolerance requirements. Conduct an appropriate hypothesis test.*  *(a) at the 5% significance level*  *(b) at the 1% significance level*  *Weights in selected apple batch (in grams): 158; 167; 149; 169; 164; 139; 154; 150; 157; 171; 152; 161; 141; 166; 172* | |
| Solution | | | The mean weight of the sample is 158 g, and the sample standard deviation is 10.4335 g. The sample size is 15.The weight tolerance (standard deviation)of the Class A apple grade is . The null hypothesis is that Class A grade apples have at most a 7.5 g weight tolerance.  *H0*: *σ 2* ≤ 7.52  *Ha*: *σ 2* > 7.52  We use a chi-square distribution with *df* = *n* – 1 = 14.  The *p*-value is  Alpha: 0.05 and 0.01  Decision: At the 5% significance level, reject the null hypothesis. At the 1% significance level, do not reject the null hypothesis.  Reason for decision: *p*-value < 0.05 but *p*-value > 0.01  Conclusion: At the 5% level, there is sufficient evidence to conclude that the weight tolerance is more than 7.5 g.  If the significance level is 0.01, there is not sufficient evidence to conclude that the weight tolerance is more than 7.5 g. | |
| Exercise 130. | | *a. Explain why a goodness-of-fit test and a test of independence are generally right-tailed tests.*  *b. If you did a left-tailed test, what would you be testing?* | |
| Solution | | a. The test statistic is always positive and if the expected and observed values are not close together, the test statistic is large and the null hypothesis will be rejected.  b. Testing to see if the data fits the distribution “too well” or is too perfect. | |

This file is copyright 2015, Rice University. All Rights Reserved.