Carrying Out a Chi-Square Test for Goodness of Fit Quiz

1. A company claims they produce their mixed bag of candies so that, of the candies in the bag, 20 percent are dark chocolate, 60 percent are milk chocolate, and 20 percent are white chocolate. In a random sample of candies of size 50, the counts are as follows: 6 dark, 32 milk, and 12 white. Assuming the conditions for inference are met, what is the test statistic for a chi-square goodness-of-fit test to investigate whether the distribution of the sample is consistent with the company's claim?

(A) $\chi^2 = 6^2 + 32^2 + 12^2$

(B) $\chi^2 = 10^2 + 30^2 + 10^2$

(C) $\chi^2 = (6-10)^2 + (32-30)^2 + (12-10)^2$

(D) $\chi^2 = \frac{(6-10)^2}{10} + \frac{(32-30)^2}{30} + \frac{(12-10)^2}{10}$ ✓

(E) $\chi^2 = \frac{(10-6)^2}{6} + \frac{(30-32)^2}{32} + \frac{(10-12)^2}{12}$

Answer D

Correct. The expected counts are $50(0.20) = 10$ for dark chocolate, $50(0.60) = 30$ for milk chocolate, and $50(0.20) = 10$ for white chocolate. The correct formula to calculate the chi-square goodness-of-fit test statistic is $\chi^2 = \Sigma \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}} = \frac{(6-10)^2}{10} + \frac{(32-30)^2}{30} + \frac{(12-10)^2}{10}$.

2. A local restaurant claims that it gets 45 percent of its customers from Monday through Thursday, 20 percent on Friday, 20 percent on Saturday, and 15 percent on Sunday. How many degrees of freedom should be used to conduct a chi-square goodness-of-fit test of the claim?

(A) 3 ✓

(B) 4

(C) 6

(D) 7

(E) It is not possible to determine the degrees of freedom without knowing the sample size.

Answer A

Correct. There are 4 categories (Monday through Thursday, Friday, Saturday, and Sunday), and the number of degrees of freedom for a chi-square goodness-of-fit test is one less than the number of categories, so $4 - 1 = 3$.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

3. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

A small coffee shop sells freshly squeezed juices in a refrigerated unit with slots where juice is displayed. These slots are called facings. The manager of the coffee shop suspects that the distribution of juice sales is different than the distribution of facings for each type of juice, so the manager records the sales of each juice over a two-week period. The proportion of facings and the sales for each type of juice are shown in the tables.

Juice	**Mango**	**Orange**	**Apple**	**Pineapple**	**Grapefruit**	**Grape**
Proportion of Facings	0.1875	0.250	0.250	0.125	0.125	0.0625

Juice	**Mango**	**Orange**	**Apple**	**Pineapple**	**Grapefruit**	**Grape**
Observed Number of Sales	23	35	46	12	10	5

(a) Construct a single bar chart that contains both the expected proportion of sales based on the proportion of facings and the observed proportion of sales for each type of juice.

(b) Assuming the conditions for inference have been met, does the coffee shop owner have sufficient evidence to conclude that the distribution of sales is proportional to the number of facings at a 5 percent level of significance? Conduct the appropriate statistical test to support your conclusion.

4-part Inference scoring

The primary goals of this question are to assess a student's ability to (1) construct a frequency bar chart; (2) identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question.

Scoring

Part (b) has three scoring steps. Those scoring steps 1, 2, 3 and part (a) are each scored as essentially correct (E), partially correct (P), or incorrect (I). So, part (b) is worth 3 Es and part (a) is worth 1 E.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as ½ point.

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· All notation is correct and clearly marked

· All explanations are clear

· No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

· No minor calculation errors are made, if they are not part of the scoring

· Interpretation parts are especially strong

Reasons to score down:

· Notation is not wrong, but is spotty and not clearly marked

· Explanations are not wrong, but are hard to follow

· Wrong or extraneous information is included but not part of scoring

· Minor calculation errors that are not part of the scoring are made

· Interpretation parts are scored an E but are considered a weak E

0	1	2	3	4 ✓

Scoring steps 1, 2, 3 and part (a) sum to 4 points

OR

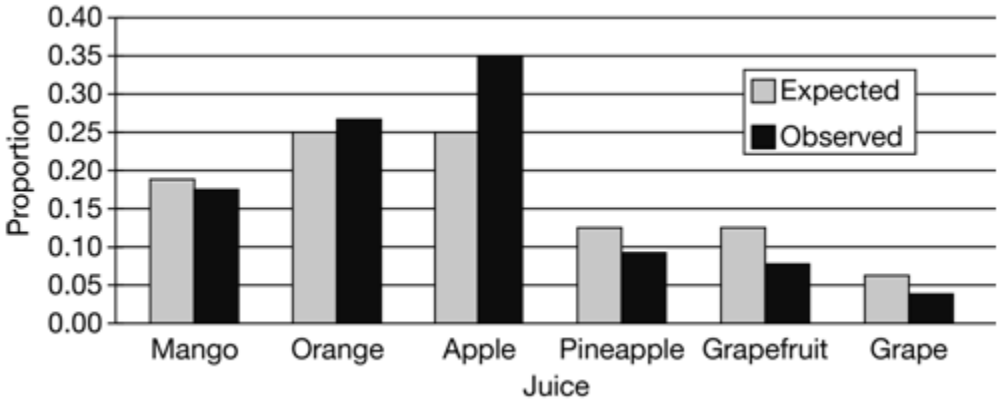
Scoring steps 1, 2, 3 and part 3½ points AND a holistic approach is used to decide to score up

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Scoring Step 1 Part (b) essentially correct
- ☐ Scoring Step 1 Part (b) partially correct
- ☐ Scoring Step 1 Part (b) incorrect
- ☐ Scoring Step 2 Part (b) essentially correct
- ☐ Scoring Step 2 Part (b) partially correct
- ☐ Scoring Step 2 Part (b) incorrect
- ☐ Scoring Step 3 Part (b) essentially correct
- ☐ Scoring Step 3 Part (b) partially correct
- ☐ Scoring Step 3 Part (b) incorrect

Solution

Part (a):

Carrying Out a Chi-Square Test for Goodness of Fit Quiz



Scoring

Part (a) is scored as follows:

Essentially correct (E) if the response correctly constructs AND labels a bar chart that contains both the expected and observed proportions

Partially correct (P) if the response fails to accurately construct a bar chart or uses frequencies instead of proportions in the plot

Incorrect (I) if the response does not satisfy the criteria for E or P

Solution

Part (b)

Scoring step 1: Identification and hypotheses

The appropriate test is a chi-square test for goodness-of-fit.

The hypotheses are:

H_0 : The distribution of juice sales is the same as the distribution of facings

H_a : The distribution of juice sales is not the same as the distribution of facings

OR

H_0 :
$p_{Mango} = 0.1875,\ p_{Orange} = 0.250,\ p_{Apple} = 0.250,\ p_{Pineapple} = 0.125,\ p_{Grapefruit} = 0.125,\ p_{Grape} = 0.0625$

H_a : At least one of the proportions is not as specified.

Scoring

Part (b)

Scoring step (1) is scored as follows:

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

Essentially correct (E) if the response satisfies the following two components:

· Correctly states, either in words or in a formula, the chi-square goodness-of-fit test

· Correct hypotheses in context

Partially correct (P) if the response includes only one of the two components

Incorrect (I) if the response does not meet the criteria for E or P

Notes:

· A response that indicates the use of $\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$ anywhere in the student's response satisfies component 1.

· The following hypotheses satisfy component 2.

H_0 : The number of juice sales is the same as the number of product facings

H_a : The number of juice sales is not the same as the number of product facings

Solution

Part (b)

Scoring step 2: Calculation

	Observed salescounts	Observed sales proportions	Proportions of facings	Expected counts
Mango	23	$\frac{23}{131} \approx 0.176$	0.1875	$(0.1875)(131) \approx 24.563$
Orange	35	$\frac{35}{131} \approx 0.267$	0.250	$(0.250)(131) \approx 32.750$
Apple	46	$\frac{46}{131} \approx 0.351$	0.250	$(0.250)(131) \approx 32.750$
Pineapple	12	$\frac{12}{131} \approx 0.092$	0.125	$(0.125)(131) \approx 16.375$
Grapefruit	10	$\frac{10}{131} \approx 0.076$	0.125	$(0.125)(131) \approx 16.375$
Grape	5	$\frac{5}{131} \approx 0.038$	0.0625	$(0.0625)(131) \approx 8.188$

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

$= \frac{(23-24.563)^{2.}}{24.563} + \frac{(35-32.75)^{2.}}{32.75} + \frac{(46-32.75)^{2.}}{32.75} + \frac{(12-16.375)^{2.}}{16.375} + \frac{(10-16.375)^{2.}}{16.375} + \frac{(5-8.188)^{2.}}{8.188}$

≈ 10.50

The value of the chi-square test statistic is approximately 10.50, with a corresponding *p*-value of approximately 0.062.

Scoring

Part (b)

Scoring step (2) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

· The correct test statistic

· The correct *p*-value *or* correct critical value (11.07)

Partially correct (P) if the response satisfies only one of the two components

Incorrect (I) if the response does not meet the criteria for E or P

Note: The *p*-value in component 2 is satisfied if it follows correctly from an incorrectly calculated value of the test statistic.

Solution

Part (b)

Scoring step 3: Conclusion with justification

Since the *p*-value is greater than the level of significance, there is not sufficient evidence to reject the null hypothesis. This means that it cannot be concluded that the observed proportion of juice sales is different from the expected number of sales based on the number of facings.

Scoring

Part (b)

Scoring step (3) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

· Comparison of *p*-value to alpha (or test statistic to critical value, i.e., the chi-square test statistic (10.50) is less than the critical value (11.07))

· Correct decision to fail to reject the null

· States conclusion in context

Partially correct (P) if the response satisfies only two of the three components

Incorrect (I) if the response does not meet the criteria for E or P

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

4. A chi-square goodness-of-fit test where all assumptions were met yielded the test statistic $\chi^2 = 12.4$. Henry claims the corresponding p–value of 0.03 means that the probability of observing a test statistic of $\chi^2 = 12.4$ is 0.03, assuming the null hypothesis is true. Which of the following is a valid criticism of this interpretation of the p–value?

(A) The null hypothesis can never be assumed to be true.

(B) The null hypothesis is not stated.

(C) The p–value is not the probability of observing 12.4 exactly. ✓

(D) The significance level is not stated.

(E) The degrees of freedom are not stated.

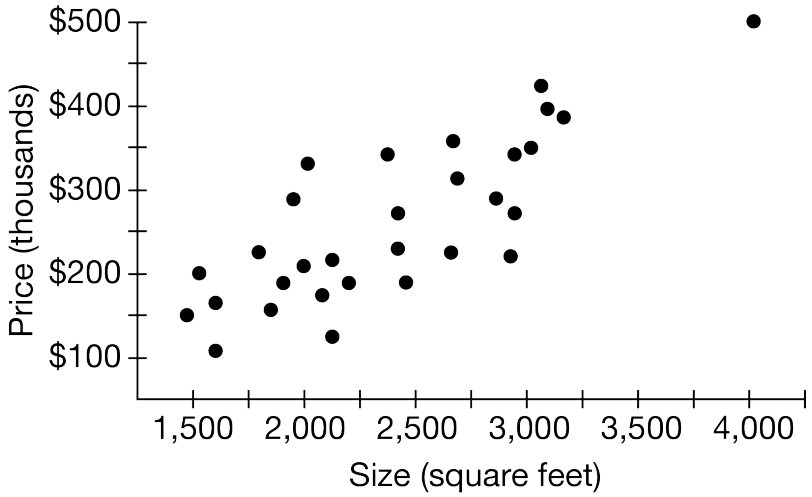
Answer C

Correct. Assuming the null hypothesis is true, the p–value is the probability of obtaining a test statistic as extreme or more extreme than the observed value, not equal to the value.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

5. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

The following scatterplot shows the size, in square feet, and the selling price, in thousands of dollars, for a sample of 30 houses for sale in a certain area.



(a) Describe the association shown in the scatterplot.

(b) In the region, houses are considered large if they are greater than 2,500 square feet and expensive if the selling price is greater than $300,000. The following two-way table summarizes the houses in the sample.

	Large	Not Large	Total
Expensive	8	2	10
Not Expensive	4	16	20
Total	12	18	30

(i) Use the information in the table to construct a graphical display of the data.

(ii) Assume there is no association between size (large, not large) and price (expensive, not expensive). Use the given totals to complete the following table with the expected number of houses for each classification if there was no association.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

	Large	Not Large	Total
Expensive			10
Not Expensive			20
Total	12	18	30

For associations displayed in the scatterplot, the strength of linear association is measured by the correlation coefficient. For the scatterplot of houses, $r = 0.82$.

For associations that are summarized in two-way tables, the strength of association is measured by the chi-square statistic. The formula for the chi-square statistic is $\chi^2 = \Sigma \frac{(observed-expected)^2}{expected}$, where expected is the count assuming no association and observed is the count shown by the data. Greater values of χ^2 indicate stronger association. For the table of counts in part (b), $\chi^2 = 10$.

(c) Suppose the selling price for the most expensive house in the sample is decreased from $\$489{,}000$ to $\$325{,}000$.

(i) What effect would the decrease have on the value of r? Explain your reasoning.

(ii) What effect would the decrease have on the value of χ^2? Explain your reasoning.

(d) Based on your answer to part (c), explain one benefit and one drawback of using χ^2 rather than r to measure the strength of an association.

Parts A, B, C, and D

Intent of Question

The primary goals of this question are to assess a student's ability to (1) describe an association shown in a scatterplot; (2) construct a graph to display the association between two categorical variables; (3) calculate expected counts when two categorical variables have no association; (4) describe the effects on two different measures of association when a value is changed; and (5) compare two different ways to measure the strength of an association.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as ½ point.

Scoring

Parts (a), (b), (c) and (d) are scored as essentially correct (E), partially correct (P), or incorrect (I).

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· *All notation is correct and clearly marked*

· *All explanations are clear*

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

· No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)

· No minor calculation errors are made, if they are not part of the scoring

· Interpretation parts are especially strong

Reasons to score down:

· Notation is not wrong, but is spotty and not clearly marked

· Explanations are not wrong, but are hard to follow

· Wrong or extraneous information is included but not part of scoring

· Minor calculation errors that are not part of the scoring are made

Interpretation parts are scored an E but are considered a weak E

0	1	2	3	4 ✓

Parts (a) through (d) sum to 4 points.

OR

Parts (a) through (d) sum to $3\frac{1}{2}$ points AND a holistic approach is used to decide to score up.

- [ ] Part (a) essentially correct
- [ ] Part (a) partially correct
- [ ] Part (a) incorrect
- [ ] Part (b) essentially correct
- [ ] Part (b) partially correct
- [ ] Part (b) incorrect
- [ ] Part (c) essentially correct
- [ ] Part (c) partially correct
- [ ] Part (c) incorrect
- [ ] Part (d) essentially correct
- [ ] Part (d) partially correct
- [ ] Part (d) incorrect

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

Solution

Part (a): There is a strong, positive, linear association between size and selling price, with a possible outlier near 4,000 square feet and selling price of almost $500,000.

Scoring

Part (a) is scored as follows.

Essentially correct (E) if the response includes the following four components:

· The response states that the association is positive.

· The response states that the association is strong or moderately strong.

· The response states that the association is linear.

· The response is in context by including the variable names.

Partially correct (P) if the response includes two or three of the four components.

Incorrect (I) if the response does not meet the criteria for E or P.

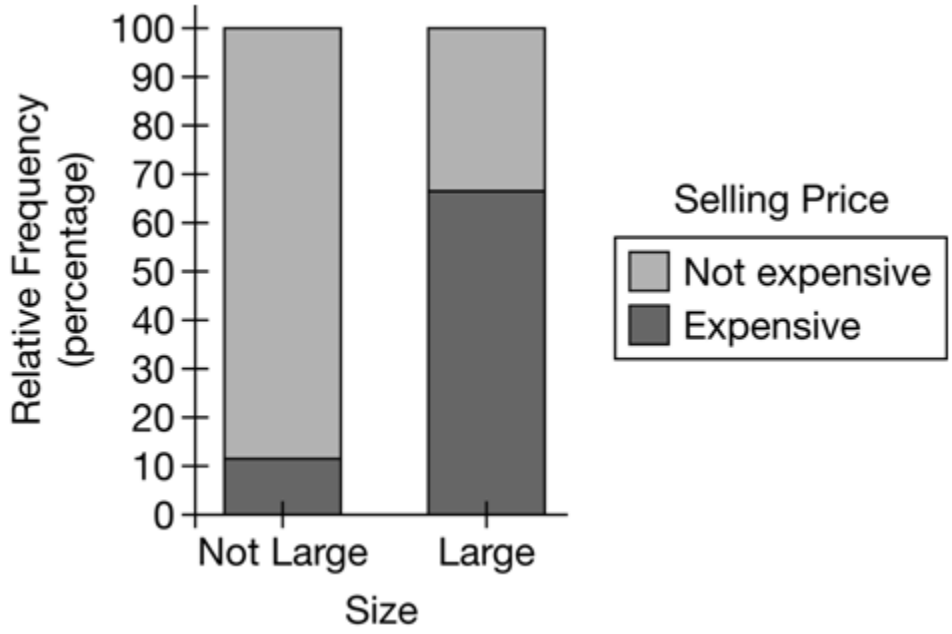
Note: The potential outlier doesn't need to be mentioned, but addressing the outlier is considered a positive for the purpose of holistic scoring.

Solution

Part (b):

Part (b-i): For the not large houses, $\frac{2}{18} \approx 11\%$ are expensive and 89% are not. For the large houses, $\frac{8}{12} \approx 67\%$ are expensive and 33% are not.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz



Part (b-ii): Overall, $\frac{10}{30} = \frac{1}{3}$ of houses are expensive. If there is no association, then $\frac{1}{3}$ of the large houses should be expensive. So, for example, $\frac{1}{3}(12) = 4$ houses should be expensive and large.

	Large	Not Large	Total
Expensive	4	6	10
Not Expensive	8	12	20
Total	12	18	30

Part (b) is scored as follows.

Essentially correct (E) if the response includes the following five components:

· In part (b-i), the graph is of an appropriate type (e.g., segmented bar graph, side-by-side bar graph) AND reasonably accurate.

· In part (b-i), the graph is well labeled.

· In part (b-i), the graph uses counts or relative frequency (percent or proportion) to compare the two groups.

· In part (b-ii), the table is completed correctly.

· In part (b-ii), the response provides appropriate work to justify the values in the table.

Partially correct (P) if the response includes three or four of the five components.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: In part (b-i), a segmented bar chart with selling price on the horizontal axis can still satisfy the first three

components if the relative frequencies within the segments are correct.

Solution

Part (c):

Part (c-i): The correlation *r* would decrease (get closer to 0) because the point would no longer be in the linear pattern of the rest of the data.

Part (c-ii): The chi-square statistic will stay the same because the house is still considered large and expensive.

Scoring

Part (c) is scored as follows.

Essentially correct (E) if the response includes the following four components:

· The response states that the correlation will decrease (get closer to 0) or get weaker.

· The response justifies why the correlation will decrease (get closer to 0) or get weaker.

· The response states that the χ^2 statistic will stay the same (equal to 10).

· The response justifies why the χ^2 statistic will stay the same (equal to 10).

Partially correct (P) if the response includes two or three of the four components.

Incorrect (I) if the response does not meet the criteria for E or P.

Solution

Part (d): A benefit of using χ^2 is that it is somewhat resistant to points that would be considered outliers on a scatterplot. However, one drawback is that all expensive/large houses are treated the same, and therefore some information is lost when measuring the strength of the relationship.

Scoring

Part (d) is scored as follows.

Essentially correct (E) if the response includes the following two components:

· The response provides an appropriate benefit of using the χ^2 statistic.

· The response provides an appropriate drawback of using the χ^2 statistic.

Partially correct (P) if the response includes one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

6. A company claims it audits its employees' transactions based on their job level. For entry-level positions, the company claims that 50 percent get a basic audit, 30 percent get an enhanced audit, and 20 percent get a complete audit. The company tests this hypothesis using a random sample and finds $\chi^2 = 0.771$ with a corresponding p–value of 0.68. Assuming conditions for inference were met, which of the following is the correct interpretation of the p–value?

(A) There is a 68 percent chance of obtaining a chi-square value of at least 0.771.

(B) There is a 68 percent chance that the company's claim is correct.

(C) If the null hypothesis were true, there would be a 68 percent chance that the company's claim is correct.

(D) If the null hypothesis were true, there would be a 68 percent chance of obtaining a chi-square value of 0.771.

(E) If the null hypothesis were true, there would be a 68 percent chance of obtaining a chi-square value of at least 0.771. ✓

Answer E

Correct. The p–value is the probability, given the null hypothesis and probability model are true, of obtaining a test statistic as extreme or more extreme than the observed value.

7. A χ^2 goodness-of-fit test was used to test the hypothesis that students at a local university select majors in the same proportions as other universities in the state. A chi-square test statistic of $\chi^2 = 45.6$ was calculated with a corresponding p-value of 0.005. Which of the following is correct?

(A) There is sufficient evidence to conclude that students at the local university do not select majors in the same proportions as do students in the rest of the state. ✓

(B) There is sufficient evidence to conclude that students at the local university select majors in the same proportions as do students in the rest of the state.

(C) There is insufficient evidence to conclude that students at the local university do not select majors in the same proportions as do students in the rest of the state.

(D) There is insufficient evidence to conclude that students at the local university select majors in the same proportions as do students in the rest of the state.

(E) Students at the local university select majors in the same proportions as do students in the rest of the state.

Answer A

Correct. There is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis since the p-value of 0.005 is very small. The null hypothesis in this case is that students at the local university select majors in the same proportions as do students at other state universities, so rejecting the null hypothesis means that students at the local university do not select majors in the same proportions as

at the other state universities.

8. A newspaper article indicated that 43 percent of cars with black seats are white, 46 percent of cars with black seats are blue, 7 percent of cars with black seats are red, and 4 percent of cars with black seats are black. A test was conducted to investigate whether the color of cars with black seats was consistent with the newspaper article. A random sample of cars of these colors was selected, and the value of the chi-square test statistic was $\chi^2 = 8.2$. Which of the following represents the p–value for the test?

(A) $P(\chi^2 \geq 8.2) = 0.08$

(B) $P(\chi^2 \geq 8.2) = 0.04$ ✓

(C) $P(\chi^2 \leq 8.2) = 0.96$

(D) $P(\chi^2 = 8.2) = 0.00$

(E) The p-value cannot be calculated because the sample size is not given.

Answer B

Correct. Since there are 4 categories, there are $4 - 1 = 3$ degrees of freedom. The area to the right of $\chi^2 = 8.2$ under the chi-square distribution with 3 degrees of freedom is represented by $P(\chi^2 \geq 8.2)$, and is approximately 0.04, as found using technology.

9. Ms. Harper knows that her students in a computing course can choose from one of three operating systems for the semester: Doors, Banana, or Duix. Ms. Harper wants to test the hypothesis that her students will select the operating systems in the same proportion as students in other computing courses at the university. She conducts a χ^2 goodness-of-fit test and calculates $\chi^2 = 3.79$ with a corresponding p-value of 0.15. Which of the following is correct at a 5-percent level of significance?

(A) Reject the null hypothesis, since $3.79 > 2$.

(B) Fail to reject the null hypothesis, since $3.79 > 2$.

(C) Reject the null hypothesis, since $0.15 > 0.05$.

(D) Fail to reject the null hypothesis, since $0.15 > 0.05$. ✓

(E) Reject the null hypothesis, since $0.15 < 3.79$.

Answer D

Correct. The p-value is correctly compared to the significance level, and the decision to fail to reject the null hypothesis is correct, since the p-value is greater than the level of significance.

Carrying Out a Chi-Square Test for Goodness of Fit Quiz

10. A certain type of legal proceeding has three possible outcomes: in favor of party A, in favor of party B, or not in favor of either party. The outcomes are expected to be 40 percent, 20 percent, and 40 percent, respectively. A random sample of 40 cases is selected from a certain judge to investigate whether the judge's outcomes are consistent with the expected outcomes. A chi-square goodness-of-fit test is conducted, and the value of the chi-square test statistic is $\chi^2 = 9.19$ with a corresponding p–value of 0.01. Assuming the conditions for inference were met, which of the following is the correct interpretation of the p–value?

(A) There is a 1 percent chance that the company's claim is correct.

(B) If the null hypothesis is true, there is a 1 percent chance that the company's claim is correct.

(C) If the null hypothesis is true, there is a 1 percent chance of obtaining a chi-square value of 9.19.

(D) If the null hypothesis is true, there is a 1 percent chance of obtaining a chi-square value of at least 9.19. ✓

(E) There is a 1 percent chance of obtaining a chi-square value of at least 9.19.

Answer D

Correct. The p–value is the probability, given the null hypothesis and probability model are true, of obtaining a test statistic as extreme, or more extreme, than the observed value.

11. A regional highway uses 8 tollbooths that are open to all vehicles. A chi-square goodness-of-fit test using a significance level of $\alpha = 0.05$ was conducted to determine whether the tollbooths are all used in equal proportions. A chi-square value of $\chi^2 = 19.1$ was calculated with a corresponding p-value of 0.008. Which of the following is correct?

(A) There is insufficient evidence to suggest that the tollbooths are not used in equal proportions.

(B) There is insufficient evidence to suggest that the tollbooths are used in equal proportions.

(C) There is sufficient evidence to suggest that the tollbooths are not used in equal proportions. ✓

(D) There is sufficient evidence to suggest that the tollbooths are used in equal proportions.

(E) The tollbooths are used in equal proportions.

Answer C

Correct. The null hypothesis in this case is that the tollbooths are used in equal proportions. The p-value is less than the significance level, so there is evidence to reject the null hypothesis in favor of the alternative hypothesis (that the tollbooths are not used in equal proportions).