Unit 8 Progress Check: MCQ Part A

1. Which of the following describes a scenario in which a chi-square goodness-of-fit test would be an appropriate procedure to justify the claim?

(A) A statistician would like to show that one geographical location has a higher proportion of dogs that shed than another geographical location has. The statistician has two independent random samples of dogs from two different geographical locations and has recorded the proportion of dogs that shed in each sample.

(B) A principal would like to investigate whether more than 50% of the students in a local high school eat in the school cafeteria. The principal has a random sample of individuals within the school and records the proportion of the students who eat lunch in the school cafeteria.

(C) A campaign manager would like to show that the distribution of individuals within several social economic categories is different than what a newspaper reported. The campaign manager has a random sample of potential voters in a large city and records the number of individuals within each of the categories.

(D) A manager of a water treatment plant would like to investigate whether there is a relationship between the amount of chemical used and the number of bacteria present in the water treated at the plant. The manager measures the level of bacteria from tanks at the facility that each received a different level of chemical treatment.

(E) City officials would like to estimate the average price of gas in their city. The officials have a random sample of gas prices at several gas stations within their city limits.

Answer C

Correct. A chi-square goodness-of-fit test would be an appropriate procedure to justify the claim. Since the campaign manager is testing to see whether the distribution of individuals within each category in the sample is different from the distribution of individuals reported in the newspaper, a chi-square goodness-of-fit test should be used.

2. An amusement park keeps track of the percentage of individuals with season passes according to age category. An independent tourist company would like to show that this distribution of age category for individuals buying season passes is different from what the amusement park claims. The tourist company randomly sampled 200 individuals entering the park with a season pass and recorded the number of individuals within each age category.

Age Category	**Child (under 13 years old)**	**Teen (13 to 19 years old)**	**Adult (20 to 55 years old)**	**Senior (56 years old and over)**
Number of Individuals	56	86	44	14

The tourist company will use the data to test the amusement park’s claim, which is reflected in the following null hypothesis. $\mathrm{H}_0 : p_{\text{child}} = 0.23$, $p_{\text{teen}} = 0.45$, $p_{\text{adult}} = 0.20$, and $p_{\text{senior}} = 0.12$. What inference procedure will the company use to investigate whether or not the distribution of age category for individuals with season passes is different from what the amusement park claims?

(A) A one-sample z-test for a population proportion

(B) A two-sample z-test for a difference between population proportions

(C) A matched pairs t-test for a mean difference

(D) A chi-square test for homogeneity

(E) A chi-square goodness-of-fit test

Answer E

Correct. A chi-square goodness-of-fit test is used to investigate whether or not there is a significant difference between a distribution generated from a sample and the hypothesized population distribution.

3. A recent article published in *Berry Weekly* reported a probability distribution for the different types of jelly that individuals prefer. Editors from a competitive magazine, *Jammin*, conducted their own study to test the distribution. The editors from *Jammin* surveyed a random sample of 50 individuals and recorded the observed counts of individuals for each jelly type. They decided to test *Berry Weekly*'s claim using a chi-square goodness-of-fit test using *Jammin*'s observed counts compared with the number of expected counts based on the *Berry Weekly* data.

	***Berry Weekly* Expected Counts**	***Jammin* Observed Counts**
Strawberry (S)	16.5	18
Grape (G)	11	12
Wild Berry (WB)	9.5	8
Peach (P)	7.5	6
Other (O)	5.5	6

Which of the following is the correct null hypothesis for the test?

(A) $\mathrm{H}_0 : p_S = 0.165, p_G = 0.11, p_{WB} = 0.095, p_p = 0.075, p_O = 0.055$

(B) $\mathrm{H}_0 : p_S = 0.33, p_G = 0.22, p_{WB} = 0.19, p_p = 0.15, p_O = 0.11$

(C) H_0 : At least one of the proportions is different.

(D) $\mathrm{H}_0 : p_S = 0.18, p_G = 0.12, p_{WB} = 0.08, p_p = 0.06, p_O = 0.06$

(E) $\mathrm{H}_0 : p_S = 0.36, p_G = 0.24, p_{WB} = 0.16, p_p = 0.12, p_O = 0.12$

Unit 8 Progress Check: MCQ Part A

Answer B

Correct. The sample size is 50, so the proportions are calculated by dividing the expected count by the sample size $\left(\text{i.e.}, p = \frac{16.5}{50} = 0.33, \text{etc.}\right)$.

4. The table displays the distribution of the percentage of different types of home heating sources for a large mountain city, as reported by the city newspaper.

Type of Heating Source	**Wood Stove**	**Electric**	**Propane/Gas**	**Solar Radiant Floor**
Percent	38%	26%	20%	16%

A chi-square goodness-of-fit test will be performed using a simple random sample of 100 homes to investigate whether the proportion of homes heated with each source is the same as what is reported by the newspaper. Which of the following represents the alternative hypothesis of the test?

(A) The proportions for the different heating systems match those reported by the newspaper.

(B) At least one of the heating source proportions is the same as the corresponding proportion reported by the newspaper.

(C) The heating sources are not evenly distributed between homes.

(D) At least one of the heating source proportions is different from the proportion reported by the newspaper.

(E) Wood stove heating represents the highest proportion of heating source.

Answer D

Correct. If it can be shown that at least one of the proportions is different, then it can be concluded that the distribution is different from the distribution that the newspaper reported.

5. A Labrador retriever club has 130 members: 65 black Labs, 44 golden Labs, and 21 chocolate Labs. Pablo is going to perform a chi-square goodness-of-fit test to see if the distribution of Labrador retrievers in the club is the same as the distribution nationally. Pablo is going to test his sample against the following null hypothesis, which reflects the national distribution.

$\text{H}_0: \; p_{black} = 0.53, p_{golden} = 0.39, p_{chocolate} = 0.08$

If the distribution of Labrador retrievers in the club were to match that of the national average, how many of each type of Labrador retriever would Pablo expect to see in his club?

Unit 8 Progress Check: MCQ Part A

(A)

Labrador Type	**Black**	**Golden**	**Chocolate**
Expected Counts	68.9	50.7	10.4

(B)

Labrador Type	**Black**	**Golden**	**Chocolate**
Expected Counts	50	33.8	16.2

(C)

Labrador Type	**Black**	**Golden**	**Chocolate**
Expected Counts	65	44	21

(D)

Labrador Type	**Black**	**Golden**	**Chocolate**
Expected Counts	43.3	43.3	43.3

(E)

Labrador Type	**Black**	**Golden**	**Chocolate**
Expected Counts	69	51	10

Answer A

Correct. Each proportion in the null hypothesis, which is the hypothesized value of the population proportion, is multiplied by the sample size (130) to determine the expected counts.

Unit 8 Progress Check: MCQ Part A

6. Jana, a high school principal, hosted a movie event at her school. Jana's assistant kept track of the number of students in each grade who attended the event. The distribution shown in the table represents the number of students in each grade that were present.

Grade Level	**Freshman**	**Sophomore**	**Junior**	**Senior**
Number	52	56	60	70

Jana knows that the grade levels are equally distributed across the school of 1,200 students. She would like to use a chi-square test to see if the proportion of individuals in each class at the movie are also equally distributed. How many seniors would be expected at the event?

(A) 840

(B) 352.9

(C) 300

(D) 70

(E) 59.5

Answer E

Correct. If the grade levels have the same proportion, then the expected proportion of seniors is 0.25. The sample size for students at the movie is $52+56+60+70=238$. Thus, the expected count of seniors is the sample size times the expected proportion of seniors, so $238(0.25)=59.5$.

7. A bag of candy contains 5 different types of colored candies; red, green, blue, yellow, and orange. According to the manufacturer, bags should contain an equal number of each color. Students in a statistics class decided to use a chi-square procedure to test the manufacturer's claim. They opened a bag of candy and recorded the number of candies of each color. The results are shown in the following table.

Red	**Green**	**Blue**	**Yellow**	**Orange**
17	24	20	25	14

Which color contributes most to the chi-square test statistic?

(A) Red

(B) Green

(C) Blue

(D) Yellow

(E) Orange

Unit 8 Progress Check: MCQ Part A

Answer E

Correct. There are 100 candies in total, so if there are an equal number of each color, the expected counts are 20 for each color. Since the count for orange, 14, is the furthest from 20, it contributes the most to the chi-square test statistic.

8. Which of the following best describes the shape of the chi-square distribution when the degrees of freedom are less than 10 ?

(A) Unimodal and symmetric

(B) Skewed to the right ✓

(C) Skewed to the left

(D) Uniform

(E) Bimodal

Answer B

Correct. The distribution would be skewed to the right.

9. A national publication showed the following distribution of favorite class subjects for high school students.

Class Subject	**Math**	**English**	**Social Studies**	**Physical Education**	**Music**	**Other**
Percentage	5%	14%	28%	26%	20%	7%

Pasquale, a student from a high school of 1,200 students, wants to see whether the distribution at his school matches that of the publication. He stands at the school entrance in the morning and asks the first 40 students he sees what their favorite class is. Pasquale records the following table of observed values.

Class Subject	**Math**	**English**	**Social Studies**	**Physical Education**	**Music**	**Other**
Observed	7	8	7	6	6	6

He decides to conduct a chi-square goodness-of-fit test to see whether his high school's distribution differs significantly from that of the publication. Pasquale's statistics teacher tells him that his information does not meet the conditions necessary for a goodness-of-fit test. Which condition has <u>not</u> been met?

I. Data are collected using a random sample or randomized experiment

II. $n \leq 0.10N$

III. All expected counts are greater than 5.

Unit 8 Progress Check: MCQ Part A

(A) I only

(B) II only

(C) III only

(D) I and III only

(E) I and II only

Answer D

Correct. Students were not selected randomly, and since the expected count for math is $40(0.05) = 2$, not all expected counts are greater than 5.

10. A round spinner is divided into five sections, where the sections do not have the same size. Using the measure of the interior angles, the probability of the spinner landing on any individual space on a spin is calculated and given in the table.

Section	**1**	**2**	**3**	**4**	**5**
Probability	0.40	0.20	0.20	0.15	0.05

A statistics student is asked to test the integrity of the spinner using a chi-square goodness-of-fit test. What is the minimum number of times the spinner should be spun to conduct this test?

(A) 5

(B) 25

(C) 50

(D) 100

(E) 500

Answer D

Correct. One of the conditions for using a chi-square goodness-of-fit test is that all the expected counts should be greater than 5. The least count will come from the section that has the least probability, 0.05. So n must be great enough so that $0.05n \geq 5$, which implies that $n \geq 100$.

11. A sports fan conducted a test to investigate whether male high school athletes are equally divided among football, soccer, swimming, tennis, and basketball. A sample of male high school athletes was selected and the resulting value of the chi-square test statistic was 10.65. Which of the following represents the p-value?

Unit 8 Progress Check: MCQ Part A

(A) $P(\chi^2 \geq 10.65) = 0.00$

(B) $P(\chi^2 \geq 10.65) = 0.03$

(C) $P(\chi^2 \geq 10.65) = 0.06$

(D) $P(\chi^2 \geq 10.65) = 0.94$

(E) $P(\chi^2 \geq 10.65) = 0.97$

Answer B

Correct. Since there are 5 categories, there are $5 - 1 = 4$ degrees of freedom. The probability of obtaining a chi-square test statistic greater than or equal to 10.65 with 4 degrees of freedom is approximately 0.03, as found using technology.

12. A job candidate at a large job fair can be classified as unacceptable, provisional, or acceptable. Based on past experience, a high-quality candidate is expected to get 80 percent acceptable ratings, 15 percent provisional ratings, and 5 percent unacceptable ratings. A high-quality candidate was evaluated by 100 companies and received 60 acceptable, 25 provisional, and 15 unacceptable ratings. A chi-square goodness-of-fit-test was conducted to investigate whether the evaluation of the candidate is consistent with past experience. What is the value of the chi-square test statistic and number of degrees of freedom for the test?

(A) $\chi^2 = \frac{(15-5)^2}{5} + \frac{(25-15)^2}{15} + \frac{(60-80)^2}{80}$ with 2df

(B) $\chi^2 = \frac{(15-5)^2}{5} + \frac{(25-15)^2}{15} + \frac{(60-80)^2}{80}$ with 3df

(C) $\chi^2 = \frac{(15-5)^2}{5} + \frac{(25-15)^2}{15} + \frac{(60-80)^2}{80}$ with 99df

(D) $\chi^2 = \frac{(5-15)^2}{15} + \frac{(15-25)^2}{25} + \frac{(80-60)^2}{60}$ with 2df

(E) $\chi^2 = \frac{(5-15)^2}{15} + \frac{(15-25)^2}{25} + \frac{(80-60)^2}{60}$ with 3df

Answer A

Correct. The correct formula to calculate the chi-square test statistic is $\chi^2 = \Sigma \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}} = \frac{(15-5)^2}{5} + \frac{(25-15)^2}{15} + \frac{(60-80)^2}{80}$. There are 3 categories (unacceptable, provisional, and acceptable), so the number of degrees of freedom is $3 - 1 = 2$.

13. A major credit card company is investigating whether the distribution of the number of credit cards used by its customers has changed from last year to this year. Customers are classified as using 1 card, 2 cards, or more than 2 cards. The company conducts a chi-square goodness-of-fit test to investigate whether there is a change in the distribution of number of cards used from last year to this year. The value of the chi-square test statistic was $\chi^2 = 7.82$ with a corresponding p-value of 0.02. Assuming the conditions for inference were met, which of the following is the correct interpretation of this p-value?

(A) There is a 2 percent chance that the company's claim is correct.

(B) There is a 2 percent chance of obtaining a chi-square value of at least 7.82.

(C) If the null hypothesis were true, there is a 2 percent chance of obtaining a chi-square value of at least 7.82. ✓

(D) If the null hypothesis were true, there is a 2 percent chance that the company's claim is correct.

(E) If the null hypothesis were true, there is a 2 percent chance of obtaining a chi-square value of 7.82.

Answer C

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.

14. A χ^2 goodness-of-fit test where all assumptions were met yielded the chi-square test statistic $\chi^2 = 1.92$ and a corresponding p-value of 0.75. The researcher interpreted the p-value as a 0.75 probability of observing a test statistic of $\chi^2 = 1.92$ or larger. What is wrong with the researcher's interpretation?

(A) The researcher did not state that the p-value is conditional on the null hypothesis being true. ✓

(B) The researcher interpreted the p-value as the probability of observing 1.92 exactly.

(C) The alternative hypothesis is not stated.

(D) The significance level is not stated.

(E) The degrees of freedom are not stated.

Answer A

Correct. The p-value of 0.75 is the probability of observing a test statistic of 1.92 or larger only under the assumption that the null hypothesis is true.

15. A researcher is investigating the claim that the proportion of television viewers who identify one of four shows as their favorite is the same for all four shows. A χ^2 goodness-of-fit test at a significance level of $\alpha = 0.05$ produced the test statistic $\chi^2 = 8.95$ with a corresponding p-value of 0.03. Which of the following is correct?

(A) There is sufficient evidence to reject the null hypothesis at the 0.05 level since the test statistic is greater than the p-value.

(B) There is not sufficient evidence to reject the null hypothesis at the 0.05 level since the test statistic is greater than the p-value.

(C) There is sufficient evidence to reject the null hypothesis at the 0.05 level since the p-value is less than the significance level.

(D) There is not sufficient evidence to reject the null hypothesis at the 0.05 level since the p-value is less than the significance level.

(E) There is sufficient evidence to reject the null hypothesis at the 0.05 level since the test statistic is greater than the significance level.

Answer C

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

16. A chi-square goodness-of-fit test using a significance level of $\alpha = 0.05$ was conducted to investigate whether the number of babies born in a town is uniformly distributed across the months of the year. The test produced a test statistic of $\chi^2 = 5.6$ with a corresponding p-value of 0.90. Which of the following is correct?

(A) Births are uniformly distributed across months.

(B) There is sufficient evidence to suggest that the distribution of births is not uniformly distributed across months.

(C) There is sufficient evidence to suggest that the distribution of births is uniformly distributed across months.

(D) There is insufficient evidence to suggest that the distribution of births is not uniformly distributed across months.

(E) There is insufficient evidence to suggest that the distribution of births is uniformly distributed across months.

Answer D

Correct. The null hypothesis in this case is that the distribution of births is uniformly distributed across months. The p-value is greater than the significance level, so the null hypothesis should not be rejected, meaning that there is not enough evidence to conclude the alternative hypothesis (that the distribution of births is <u>not</u> uniformly distributed across months).