Setting Up a Chi-Square Test for Homogeneity or Independence Quiz

1. In 2010, a medical research group reported the results of an experiment to evaluate the effectiveness of acupuncture to treat a chronic intestinal condition. A group of volunteers with the chronic intestinal condition agreed to participate in the experiment and be randomly assigned to either a true acupuncture treatment or a placebo treatment. The placebo treatment mimicked the application of acupuncture, but no needle penetrated the skin. Random assignment resulted in 78 subjects receiving acupuncture and 75 subjects receiving the placebo treatment. After receiving 6 treatments over the course of 3 weeks, patients were asked to report whether they had experienced a reduction in the chronic intestinal condition. The table summarizes the data from the study, with expected cell counts in parentheses.

	Yes	**No**	**Total**
Acupuncture	41 (37.2)	37 (40.8)	78
Placebo treatment	32 (35.8)	43 (39.2)	75
Total	73	80	153

Which of the following is true about the chi-square test for homogeneity?

(A) The number of subjects randomly assigned to each treatment is not the same; therefore, it is not appropriate to use a chi-square test for homogeneity across treatment groups.

(B) Volunteers do not constitute a random sample from the population of all patients with the chronic intestinal condition; therefore, it is not appropriate to use a chi-square test for homogeneity across treatment groups.

(C) Volunteers with the chronic intestinal condition were randomly assigned to each treatment, so the independence condition has been met.

(D) Not all of the observed cell counts are large enough to satisfy the conditions for applying the chi-square test of homogeneity.

(E) Not all of the expected cell counts are large enough to satisfy the conditions for applying the chi-square test for homogeneity.

2. A polling organization uses random digit dialing of registered voters in a county to gauge the voters' opinions about a ballot initiative to increase taxes that would pay for a new county park. The polling organization will investigate whether there are differences in how voters respond based on the highest level of education completed. The phone survey asked the voters contacted whether they approve or disapprove of the tax increase. The voters were also asked to indicate their highest level of education completed: high school, bachelor's degree, or master's degree.

Which of the following is the appropriate test to investigate whether there is an association between opinion about the ballot initiative and highest level of education completed?

(A) A two-sample t-test for a difference between means

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

(C) A chi-square test of homogeneity

(D) A chi-square test of independence

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

Setting Up a Chi-Square Test for Homogeneity or Independence Quiz

3. A random sample of 500 people were classified by their ages into 3 age-groups: 29 years and younger, 30 to 64 years, and 65 years and older. Each person from the sample was surveyed about which of 4 major brands of cell phone they used. Their responses were compiled and displayed in a 3-by-4 contingency table. A researcher will use the data to investigate whether there is an association between cell phone brand and age-group.

Which of the following is the appropriate test for the investigation?

(A) A one-sample t-test for a population mean

(B) A two-sample t-test for a difference between means

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

(D) A chi-square test of homogeneity

(E) A chi-square test of independence

4. A simple random sample of 1,000 United States adults was collected to investigate whether gender and being a coffee drinker are independent. The results are summarized in the table with the expected counts shown in parentheses.

		Coffee Drinker?		
		Yes	**No**	**Total**
Gender	**Male**	191 (148.5)	299 (341.5)	490
	Female	112 (154.5)	398 (355.5)	510
	Total	303	697	1,000

Which statement is true about whether the conditions for the chi-square test for independence have been met?

(A) All necessary conditions are satisfied to apply a chi-square test for independence between gender and being a coffee drinker.

(B) The data is not the result of two independent random samples; therefore, the conditions for applying the chi-square test for independence between gender and being a coffee drinker are not met.

(C) Not all of the expected cell counts are large enough to satisfy the conditions for applying the chi-square test for independence between gender and being a coffee drinker.

(D) Not all of the observed cell counts are large enough to satisfy the conditions for applying the chi-square test for independence between gender and being a coffee drinker.

(E) The total sample size is not at least 10 percent of the population of United States adults; therefore, the conditions for applying the chi-square test for independence between gender and being a coffee drinker are not met.

5. Polling organizations regularly collect data on the public's opinions and habits. A question on a recent survey asked, "How often do you purchase coffee from a coffeehouse?" There may be differences in how people respond to this question based on whether the person is a full-time student. Suppose a polling organization uses random digit dialing of local phone numbers to take a poll and asks respondents whether they are full-time students. In addition, they ask respondents to identify how often they purchase coffee from a coffeehouse (never, once a week, two to three times a week, daily, more than once a day). The data are collected in a 5-by-2 table of counts. Which of the following is the appropriate null hypothesis when conducting a chi-square test for the data?

Setting Up a Chi-Square Test for Homogeneity or Independence Quiz

(A) H_0 : The proportion of people for each category in the 5-by-2 table will be 10 percent.

(B) H_0 : For each purchasing frequency, the proportion of full-time students is different.

(C) H_0 : How often a person purchases coffee from a coffeehouse is not associated with whether a person is a full-time student.

(D) H_0 : How often a person purchases coffee from a coffeehouse is dependent on whether a person is a full-time student.

(E) H_0 : There is an association between how often a person purchases coffee from a coffeehouse and whether a person is a full-time student.

6. A reporter intends to survey residents of a city to investigate whether there are differences in use of grocery delivery services by region of the city where the residents live (north, south, west, and east). City residents will be asked to respond yes or no to the question "Do you regularly use a grocery delivery service to purchase groceries?" Results will be collected in a 4-by-2 table of counts organized by region and response to the question. For a chi-square test for homogeneity, which of the conditions listed below is <u>not</u> necessary to investigate whether there are differences between use of grocery delivery service by region?

(A) For each cell in the table, $\frac{(\text{row total})(\text{column total})}{\text{table total}}$ will be greater than 5.

(B) Data should be collected using a stratified random sample, with region as strata.

(C) All eight expected cell counts should be greater than 5.

(D) The number of residents sampled from each region should be greater than 30.

(E) The total number of residents sampled should be at most 10 percent of the total number of residents in the city.

7. Which of the following is a reason <u>not</u> to use a chi-square test of homogeneity to analyze a set of data?

(A) The data consist of one categorical variable for two or more different populations and are summarized by counts in a two-way table.

(B) The data were obtained through a simple random sample from a single population and summarized by counts on two categorical variables.

(C) The data were obtained from more than two populations to investigate whether the proportions for categorical data collected are the same.

(D) The data were obtained from four different regions to investigate whether the distribution of a categorical variable is different across the four regions.

(E) The data were obtained using a simple random sample of a population from last year and a simple random sample of the same population from this year where the data collected were categorical variables.

Setting Up a Chi-Square Test for Homogeneity or Independence Quiz

8. The ecology club at a local high school took separate random samples of sophomores, juniors, and seniors to learn whether there are differences in attitudes toward recycling by grade level. Students were asked to respond yes or no to the question, "Do you regularly recycle plastic bottles?" The results were summarized in a 3-by-2 table of counts organized by grade level and yes/no response to the recycling question. The club performed a chi-square test for homogeneity across the three grade levels. Which of the following could be an appropriate null hypothesis used by the club?

I. $H_0 : p_1 = p_2 = p_3$, where p_1 is the proportion of the sample of sophomores that responded yes, p_2 is the proportion of the sample of juniors that responded yes, and p_3 is the proportion of the sample of seniors that responded yes.

II. H_0 : There is an association between grade level and whether or not a student regularly recycles plastic bottles.

III. H_0 : There is no difference in the distribution of regular recyclers across the three grade levels.

(A) II only

(B) III only

(C) I and II only

(D) I and III only

(E) I, II, and III

9. A large factory that builds machines has three shifts, one that starts at 4:00 A.M., one that starts at noon, and one that starts at 8:00 P.M. The manager of the factory wanted to know whether there is an association between an employee's work experience (less than five years with the company, between five and twenty years with the company, over twenty years with the company) and the time of the employee's shift. The manager selected a random sample of 125 employees and classified employees by their shift time and work experience. Which of the following is an appropriate pair of hypotheses for the manager to use?

(A) H_0 : Work experience is independent of the time of an employee's shift.
H_a : Work experience is dependent on the time of an employee's shift.

(B) H_0 : Employees with over 20 years of experience are just as likely to start work at 8 P.M.
H_a : Employees with over 20 years of experience are less likely to start work at 8 P.M.

(C) H_0 : There is an association between work experience and an employee's shift.
H_a : There is no association between work experience and an employee's shift.

(D) H_0 : For each shift, there is no difference between the proportions of work experience level.
H_a : For each shift, there is a difference between the proportions of work experience level.

(E) H_0 : Among those in the sample, the proportion of employees for each work experience level did not differ by shift.
H_a : Among those in the sample, the proportion of employees for each work experience level did differ by shift.