• Baseball cards are trading cards that feature data on a player's performance in baseball games. Michelle is at a national baseball card collector's convention with approximately 20,000 attendees. She notices that some collectors have both regular cards, which are easily obtained, and rare cards, which are harder to obtain. Michelle believes that there is a relationship between the number of months a collector has been collecting baseball cards and whether the majority of the cards in their collection are regular or rare. She obtains information from a random sample of 500 baseball card collectors at the convention and records how many full months they have been collecting baseball cards and whether the majority of the cards in their card collection are regular or rare. Her results are displayed in a two-way table.

Majority Type of Baseball Cards and Months of Collecting Baseball Cards

	Fewer Than 6 Months	6 - 10 Months	11 - 15 Months	16 - 20 Months	21 or More Months	Total
Has a Majority of Regular Baseball Cards	80	84	71	76	112	423
Has a Majority of Rare Baseball Cards	11	16	9	6	35	77
Total	91	100	80	82	147	500

(A) If one collector from the sample is selected at random, what is the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of regular baseball cards? Show your work.

(B) Given that a randomly selected collector from the sample has been collecting baseball cards for fewer than 6 months, what is the probability the collector has a majority of regular baseball cards? Show your work.

(C) Michelle believes there is a relationship between the number of months spent collecting baseball cards and which type of card is the majority in the collection (regular or rare).

i. Name the hypothesis test Michelle should use to investigate her belief. Do not perform the hypothesis test.

ii. State the appropriate null and alternative hypotheses for the hypothesis test you identified in (c-i). Do not perform the hypothesis test.

(D) After completing the hypothesis test described in part (c), Michelle obtains a p-value of 0.0075. Assuming the conditions for inference are met, what conclusion should Michelle make about her belief? Justify your response.

Solution:

(A) $P(\text{11+ months} \cap \text{majority regular cards})$
$= \dfrac{71 + 76 + 112}{500}$
$= \dfrac{259}{500}$
$= 0.518$

(B) $P(\text{majority regular cards} \mid \text{fewer than 6 months})$
$= \dfrac{P(\text{majority regular cards} \cap \text{fewer than 6 months})}{P(\text{fewer than 6 months})}$
$= \dfrac{\frac{80}{500}}{\frac{91}{500}} = \dfrac{80}{91}$
≈ 0.879

(C) (i) Michelle should conduct a chi-square test for independence between months collecting baseball cards and majority card status for all baseball card collectors at the convention.

(ii) H_0: There is not an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention.
H_a: There is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention.
OR
H_0: Months collecting cards and majority card status are independent for all baseball card collectors at the convention.
H_a: Months collecting cards and majority card status are not independent for all baseball card collectors at the convention.

(D)

Because the p -value of 0.0075 is less than any reasonable α level such as 0.05 or 0.10, the null hypothesis should be rejected. The data provide convincing statistical evidence that there is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention.

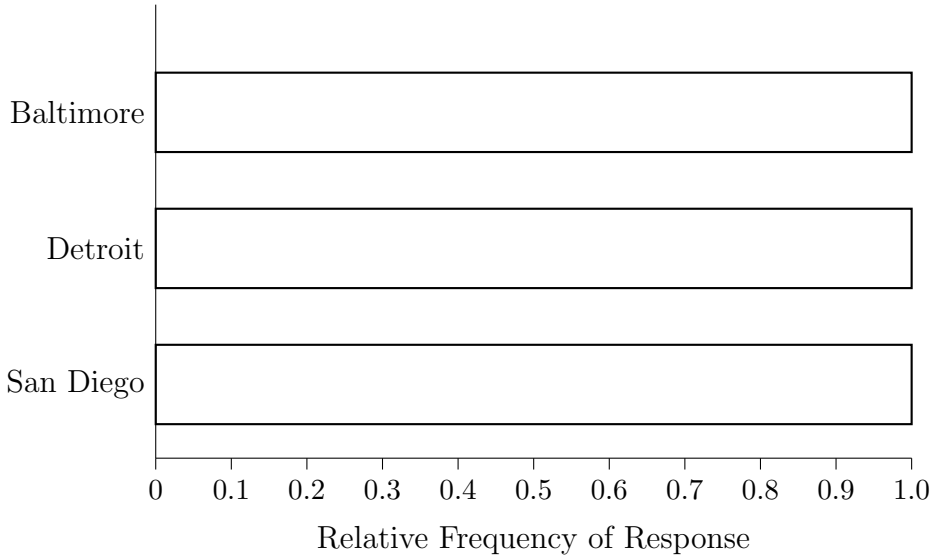
• A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

(A) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

(B) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

(C) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

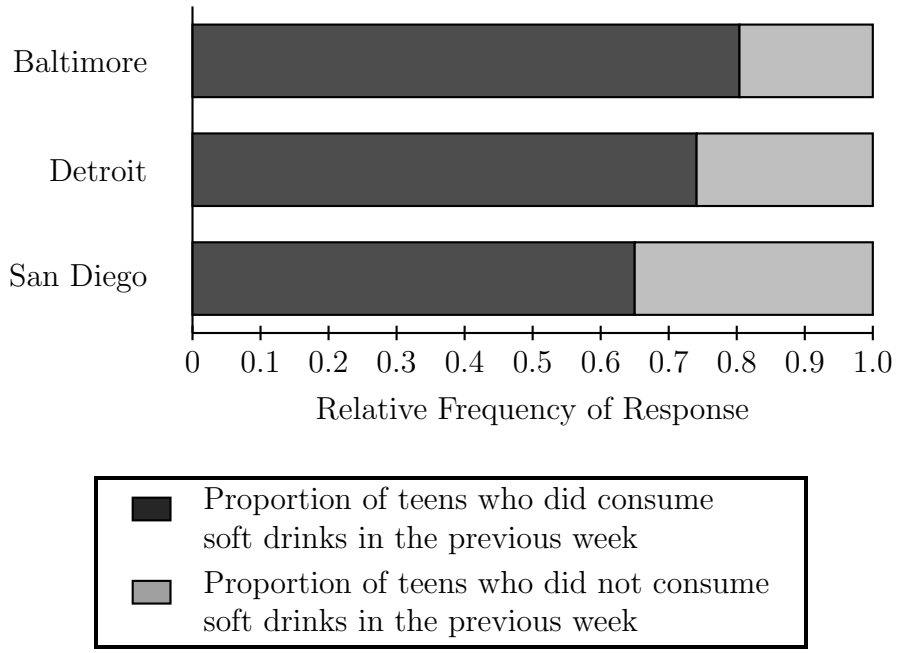
(i) Identify the appropriate inference procedure.

(ii) Identify the hypotheses of the test.

Solutions:

(A) No, the researcher's claim is not correct. Although the Baltimore survey has the least number of teens who consumed a soft drink in the past week, it also has the least number of teens surveyed among the three cities' samples. Comparing the numbers of teens who consumed a soft drink in the past week is meaningless without considering the sample sizes. The comparison should be based on proportions rather than counts. In fact, the proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportions for either of the other two cities, $\frac{1,232}{1,663} \approx 0.741$ for Detroit and $\frac{1,482}{2,280} = 0.65$ for San Diego.

(B) **(i)** A segmented bar graph of the relative frequencies based on the information in the table is shown below:



(ii) The proportion of teens who consumed a soft drink in the previous week are shown below:

- Baltimore: $\frac{727}{904} \approx 0.804$
- Detroit: $\frac{1,232}{1,663} \approx 0.741$
- San Diego: $\frac{1,482}{2,280} = 0.65$

San Diego has the smallest proportion of teens (0.65) who consumed a soft drink in the previous week.

(C)

(i) Since the data were collected from independent random samples from the three cities, a chi-square test for homogeneity should be conducted.

(ii) The appropriate hypotheses are:

H_0: There is no difference in the proportion of all teens who consumed a soft drink in the past week across the three cities.

H_a: There is at least one difference in the proportion of all teens who consumed a soft drink in the past week across the three cities.

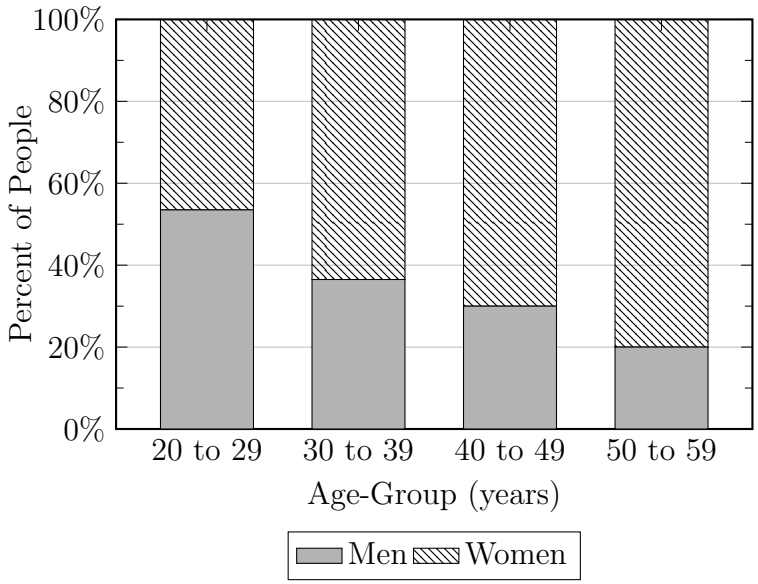
OR

H_0: The proportion of all teens who consumed a soft drink in the past week is the same across the three cities.

H_a: The proportion of all teens who consumed a soft drink in the past week differs for at least two of the three cities.

• The table and the bar chart below summarize the age at diagnosis, in years, for a random sample of 207 men and women currently being treated for schizophrenia.

	Age-Group (years)				
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

Solution

Step 1: State a correct pair of hypotheses.

The null hypothesis is that age group at diagnosis and gender are independent (that is, they are not associated) for the population of people currently being treated for schizophrenia.

The alternative hypothesis is that age group at diagnosis and gender are not independent for the population of people currently being treated for schizophrenia.

Step 2: Identify a correct test procedure (by name or formula) and check appropriate conditions.

The appropriate test is a chi-square test of independence.

The conditions for this test are satisfied because:

1. The question states that the sample was randomly selected.

2. The expected counts for the eight cells of the table are at least 5, as seen in the following table, with expected counts shown below observed counts.

	Age at Diagnosis				
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
	56.91	36.22	17.25	8.62	
Men	53	23	9	3	88
	42.09	26.78	12.75	6.38	

Step 3: Find the value of the test statistic and the p-value.

The test statistic is calculated as $\chi^2 = \sum \frac{(O-E)^2}{E}$, or

$$\begin{aligned}\chi^2 &= 2.093 + 0.395 + 0.817 + 1.322 \\ &\quad + 2.830 + 0.534 + 1.105 + 1.788 \\ &= 10.884.\end{aligned}$$

The p-value is $P(\chi^2 \geq 10.884) = 0.012$, based on $(4-1) \times (2-1) = 3$ degrees of freedom.

Step 4: State the conclusion in context, with linkage to the p-value.

Because the p-value is very small (for instance much smaller than $\alpha = 0.05$), we reject the null hypothesis and conclude that the sample data provide strong evidence that there is an association between age group at diagnosis and gender for the population currently being treated for schizophrenia.

• Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21	4
B	Apple-Zuties only	13	12
C	Neither	22	3

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

Solution

(A)

Step 1: States a correct pair of hypotheses.

H_0 : The proportion of children who would choose each snack is the same regardless of which type of ad is viewed.
H_a : The proportion of children who would choose each snack differs based on which type of ad is viewed.

Step 2: Identifies a correct test procedure (by name or formula) and checks appropriate conditions.

The appropriate procedure is a chi-square test of homogeneity.

The conditions for this test are satisfied because (1) the question states that the children were randomly assigned to groups, and (2) expected counts for the six cells of the table are all at least 5, as

seen in the following table that lists expected counts beside observed counts.

Group	Choco-Zuties	Apple-Zuties	Total
A	21 (18.67)	4 (6.33)	25
B	13 (18.67)	12 (6.33)	25
C	22 (18.67)	3 (6.33)	25
Total	56	19	75

Step 3: Calculates the appropriate test statistic and p-value.

The test statistic is calculated as $\chi^2 = \sum \frac{(O-E)^2}{E}$, which is

$$\begin{aligned}\chi^2 \approx & 0.292 + 0.860 + \\ & 1.720 + 5.070 + 0.595 + 1.754 \approx 10.291.\end{aligned}$$

The p-value is $P(\chi^2_{\text{df2}} \geq 10.291) \approx 0.006$.

Step 4: States a correct conclusion in the context of the study, using the result of the statistical test.

Because the p-value is very small (for instance, much smaller than $\alpha = 0.05$), we reject the null hypothesis at the 0.05 level (and at the 0.01 level). The data provide convincing statistical evidence that the proportions who would choose each snack differ based on which ad is viewed.

(B)

When neither ad was viewed, $\frac{22}{25}$ or 88 percent of the children chose Choco-Zuties whereas only 12 percent chose Apple-Zuties.

When the Choco-Zuties ad was viewed, 84 percent of the children chose Choco-Zuties, which was very similar to the percentage that chose them without viewing any ad. So watching the Choco-Zuties ad did not affect the snack choice very much.

When the Apple-Zuties ad was viewed, only $\frac{13}{25}$ or 52 percent of the children chose Choco-Zuties, and 48 percent chose Apple-Zuties. Watching the Apple-Zuties ad seemed to increase the proportion of children choosing Apple-Zuties.

Problems adapted from the College Board released tests.