

5. A researcher is investigating whether there is an association among professional athletes between age-group (in years) and type of sport played (basketball, football, baseball). The age-group and type of sport played for all 4,193 professional athletes in these sports for a recent year are given in the two-way table.

Age-Group by Type of Sport Played

Age (years)	Basketball	Football	Baseball	Total
Age < 25	232	807	259	1,298
$25 \leq \text{Age} < 30$	175	1,326	620	2,121
$30 \leq \text{Age} < 35$	90	287	276	653
$35 \leq \text{Age}$	19	41	61	121
Total	516	2,461	1,216	4,193

Part A

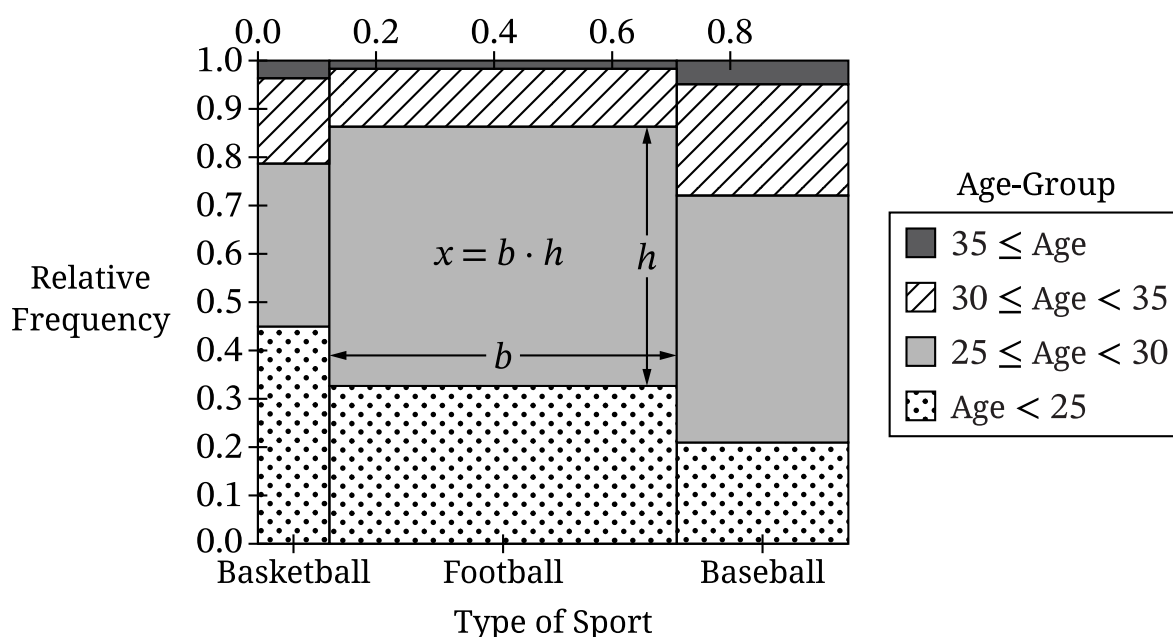
Consider the two-way table.

- What is the probability a randomly selected professional athlete is a football player?
Show your work.
- What is the probability a randomly selected professional athlete is in the age-group of " $25 \leq \text{Age} < 30$ " given they are a football player? Show your work.

Part B

A mosaic plot was constructed using the information in the two-way table. Use the mosaic plot to answer the following questions.

Relative Frequency of Age-Group by Type of Sport Played



- The b and h displayed in the mosaic plot each represent a probability calculated in part A. Which probability does b correspond to: the probability calculated in part A (i) or the probability calculated in part A (ii)?
- What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

- Are the events “Baseball” and “35 ≤ Age” mutually exclusive events? Explain.
- Are the events “Baseball” and “35 ≤ Age” independent events? Show your work.

Part D

Consider the data of all professional athletes from the two-way table. Determine if it is appropriate to carry out a chi-square test for independence to investigate whether there is an association between age-group and sport played using these data. Explain your answer.

5. 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 (cards appearing more often) 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.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

- (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.

GO ON TO THE NEXT PAGE.

5. 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

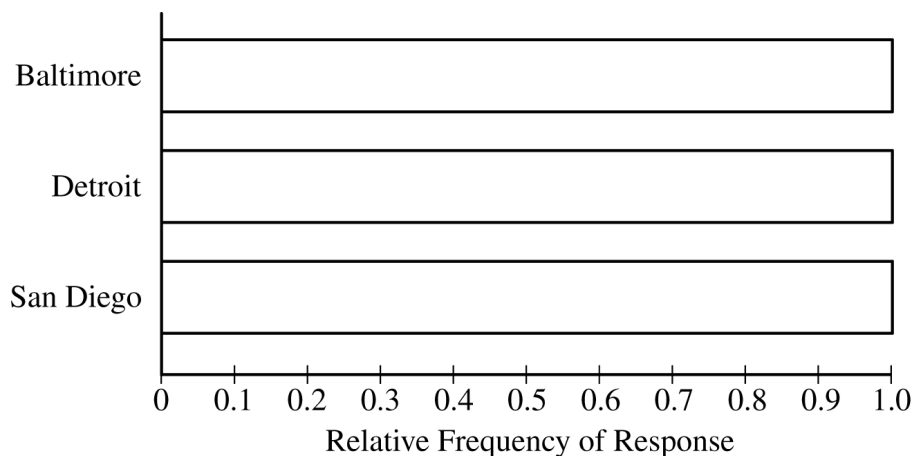
- (a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the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

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

(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.

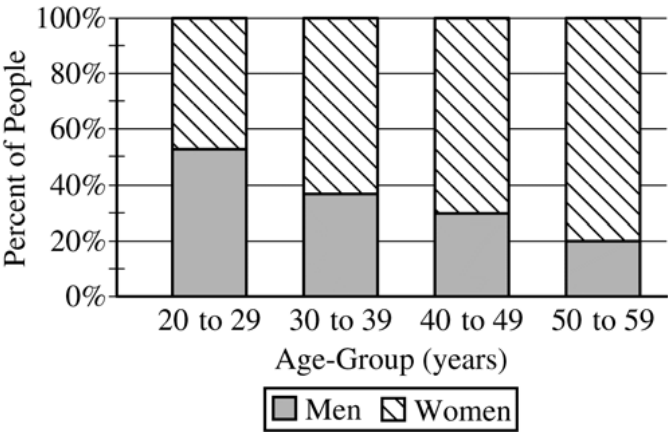
(ii) Identify the hypotheses of the test.

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

5. 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?

2. 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.