Past-paper walk-through

Statistics for Two Categorical Variables ▪ 2.3

eXercise

Questions in this set

- 2026 Q5
- 2024 Q2
- 2024 Q5
- 2022 Q6
- 2021 Q5
- 2016 Q2
- 2014 Q1

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 5

2026 ▪ Q5

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
$25 \leq$ Age < 30	175	1,326	620	2,121
$30 \leq$ Age < 35	90	287	276	653
$35 \leq$ Age	19	41	61	121
Total	516	2,461	1,216	4,193

(a)(i) Consider the two-way table.

What is the probability a randomly selected professional athlete is a football player? Show your work.

Question 5

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

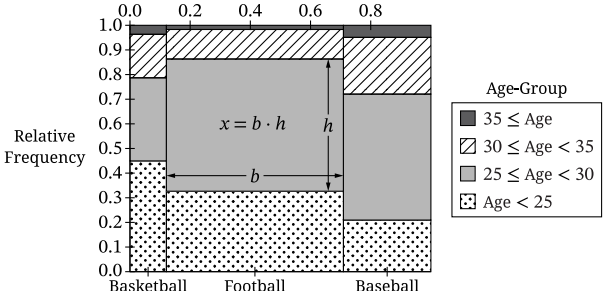
Question 5

ii. 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.

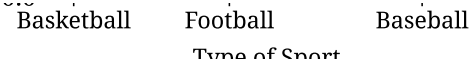
Part B

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



Question 5



Question 5

Age (years)	Basketball	Football	Baseball	Total
$\text{Age} < 25$	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

Question 5

2026 ▪ Q5

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
$25 \leq$ Age < 30	175	1,326	620	2,121
$30 \leq$ Age < 35	90	287	276	653
$35 \leq$ Age	19	41	61	121
Total	516	2,461	1,216	4,193

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

[Mosaic plot titled "Relative Frequency of Age-Group by Type of Sport Played"; horizontal axis "Type of Sport" with three columns Basketball, Football, Baseball

Question 5

whose widths are proportional to each sport's share of the total 4,193 athletes; vertical axis "Relative Frequency" from 0.0 to 1.0. Each column is divided into four stacked segments by Age-Group, shaded by the legend: dotted = Age < 25, light gray = $25 \leq$ Age < 30, hatched = $30 \leq$ Age < 35, dark gray = $35 \leq$ Age. The Football column is labeled with b as its width (horizontal extent) and h as the height of one of its segments, with the relation $x = b \cdot h$ shown inside a rectangle of that column, where x denotes the area of that rectangle.]

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)?

(b)(ii) What probability does the x displayed in the mosaic plot represent in context?

Question 5

2026 ▪ Q5

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
$25 \leq$ Age < 30	175	1,326	620	2,121
$30 \leq$ Age < 35	90	287	276	653
$35 \leq$ Age	19	41	61	121
Total	516	2,461	1,216	4,193

(c)(i) Use the information to answer the following questions.

Are the events "Baseball" and "$35 \leq$ Age" mutually exclusive events? Explain.

(c)(ii) Are the events "Baseball" and "$35 \leq$ Age" independent events? Show your work.

Question 5

2026 ▪ Q5

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
$25 \leq$ Age < 30	175	1,326	620	2,121
$30 \leq$ Age < 35	90	287	276	653
$35 \leq$ Age	19	41	61	121
Total	516	2,461	1,216	4,193

(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.

Question 2

2024 ▪ Q2

A local elementary school decided to sell bottles printed with the school district's logo as a fund-raiser. The students in the elementary school were asked to sell bottles in three different sizes (small, medium, and large). The relative frequencies of the number of bottles sold for each size by the elementary school were 0.5 for small bottles, 0.3 for medium bottles, and 0.2 for large bottles.

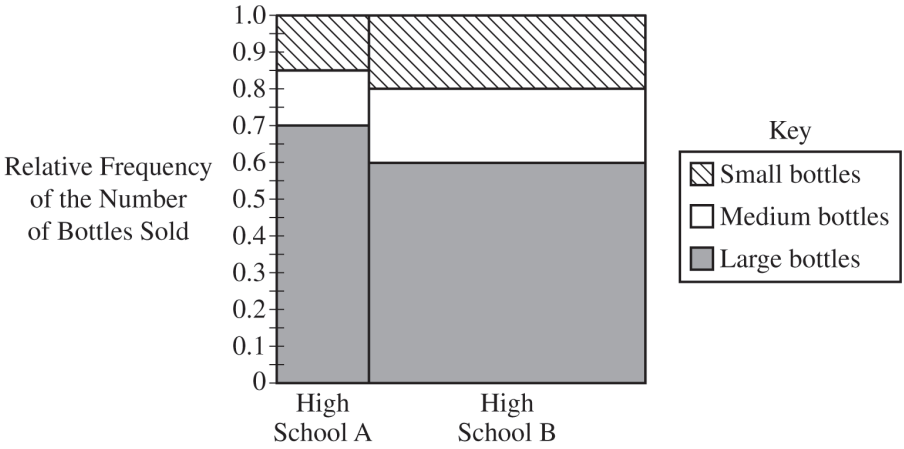
A local middle school also decided to sell bottles as a fund-raiser, using the same three sizes (small, medium, and large). The middle school students sold three times the number of bottles that the elementary school students sold. For the middle school students, the proportion of bottles sold was equal for all three sizes.

(a) Complete the segmented bar graphs representing the relative frequencies of the number of bottles sold for each size by students at each school.

Question 2

[Two blank segmented (stacked) bar graphs to be completed by the student, each a single horizontal bar on an axis "Relative Frequency of Number of Bottles Sold" from 0 to 1.0 (gridlines at 0.2, 0.4, 0.6, 0.8, 1.0): one labeled "Elementary School", one labeled "Middle School". A key at the right shows three segment patterns: hatched/cross-hatched = Small bottles, blank/white = Medium bottles, gray/shaded = Large bottles.]

Question 2


Relative Frequency of the Number of Bottles Sold
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
High School A
High School B
Key
Small bottles
Medium bottles
Large bottles

Question 2

2024 ▪ Q2

A local elementary school decided to sell bottles printed with the school district's logo as a fund-raiser. The students in the elementary school were asked to sell bottles in three different sizes (small, medium, and large). The relative frequencies of the number of bottles sold for each size by the elementary school were 0.5 for small bottles, 0.3 for medium bottles, and 0.2 for large bottles.

A local middle school also decided to sell bottles as a fund-raiser, using the same three sizes (small, medium, and large). The middle school students sold three times the number of bottles that the elementary school students sold. For the middle school students, the proportion of bottles sold was equal for all three sizes.

(b) An administrator at the elementary school concluded that the elementary school students sold more small bottles than the middle school students did. Is the elementary school administrator's conclusion correct? Explain your response.

Question 2

2024 ▪ Q2

A local elementary school decided to sell bottles printed with the school district's logo as a fund-raiser. The students in the elementary school were asked to sell bottles in three different sizes (small, medium, and large). The relative frequencies of the number of bottles sold for each size by the elementary school were 0.5 for small bottles, 0.3 for medium bottles, and 0.2 for large bottles.

A local middle school also decided to sell bottles as a fund-raiser, using the same three sizes (small, medium, and large). The middle school students sold three times the number of bottles that the elementary school students sold. For the middle school students, the proportion of bottles sold was equal for all three sizes.

(c)(i) Two high schools are also selling the bottles and are competing to see which one sold more large bottles.

Question 2

A mosaic plot for the distribution of the number of bottles sold by each of the high schools is shown here.

[Mosaic plot titled "Distribution of the Number of Bottles Sold by High School"; y-axis "Relative Frequency of the Number of Bottles Sold" from 0 to 1.0 in increments of 0.1; two columns, "High School A" and "High School B", with widths proportional to the total number of bottles sold (High School A narrower, High School B wider). Key: hatched = Small bottles, blank/white = Medium bottles, gray/shaded = Large bottles. For High School A (reading bottom to top): Large bottles (gray) from 0 to about 0.7; Medium bottles (blank) from about 0.7 to 0.85; Small bottles (hatched) from about 0.85 to 1.0. For High School B (reading bottom to top): Large bottles (gray) from 0 to 0.6; Medium bottles (blank) from 0.6 to 0.8; Small bottles (hatched) from 0.8 to 1.0.]

Question 2

Which of the two high schools sold a greater proportion of large bottles? Justify your answer.

(c)(ii) Which of the two high schools sold a greater number of large bottles? Justify your answer.

Solution 2

(a) Mark scheme

- Complete each bar so the segments' widths equal the actual relative frequencies for that school: the Elementary School bar is partitioned into small/medium/large sections matching the elementary data (using the given hatched = small, blank = medium, gray = large key), and the Middle School bar is partitioned into three EQUAL-width sections (each representing a relative frequency of about $1/3 \approx 0.33$) for small/medium/large, since the middle school sold the same number of each size. Full credit requires both bars correctly partitioned – the elementary bar per its true proportions AND the middle-school bar into three equal areas. The order of segments (e.g. small-medium-large vs. large-medium-small) does not

Solution 2

matter, a response may supply its own key defining small/medium/large instead of the given one, and cutoffs within ± 0.05 of the actual value are acceptable.

Solution 2

(b) Mark scheme

- No – the elementary administrator's conclusion is not correct. The segment for small bottles is wider (a greater relative frequency) for the elementary school than for the middle school, but the middle school sold three times as many total bottles as the elementary school. If the elementary school sold x bottles total, it sold $0.5x$ small bottles; the middle school sold $3x$ bottles total, of which $0.\overline{3}(3x) = x$ were small. Example: if the elementary school sold 100 bottles total, it sold $0.5(100) = 50$ small bottles, while the middle school sold 300 bottles total, of which $(0.\overline{3})(300) = 100$ were small; since $100 > 50$, the middle school actually sold MORE small bottles in absolute count even though its proportion of small bottles is lower. Full credit needs: (1) states the elementary administrator is

Solution 2

incorrect, (2) correct mathematical support verifying the middle school sold more small bottles (consistent with part (a)), (3) context – the words "elementary," "middle," and "bottles" must appear.

Solution 2

(c)(i)

Mark scheme

- High School A sold the greater PROPORTION of large bottles. From the mosaic plot, the relative frequency (height) of the large-bottles (gray) segment is 0.7 for High School A and 0.6 for High School B, and $0.7 > 0.6$. Full credit needs the correct school identified (A) AND the justification based on the height/relative frequency of the large-bottle segments (if the response refers to "proportions" rather than relative frequencies, it must give the values 0.7 and 0.6).

Solution 2

(c)(ii) Mark scheme

- High School B sold the greater NUMBER of large bottles. The number of bottles sold is represented by the AREA of the shaded region, not its height/proportion – High School B's column is wider (more total bottles sold) than High School A's, so even though a smaller share (0.6 vs 0.7) of High School B's bottles were large, the area of the rectangle representing large bottles sold by High School B is clearly larger than the area for High School A. Full credit needs High School B identified AND the justification based on comparing the AREAS of the large-bottle rectangles (reasonable estimates of the areas are acceptable).

Question 5

2024 ▪ Q5

Question 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.

Question 5

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

Question 5

(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.

Question 5

2024 ▪ Q5