

Past-paper walk-through

Setting Up a Chi-Square Test for Homogeneity or Independence ■ 8.5

eXercise

Questions in this set

- 2026 Q5
- 2024 Q5
- 2021 Q5
- 2017 Q5
- 2016 Q2

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 ≤ Age < 30	175	1,326	620	2,121		30 ≤ Age < 35
90	287	276	653		35 ≤ Age	19	41	61	121		Total
4,193											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

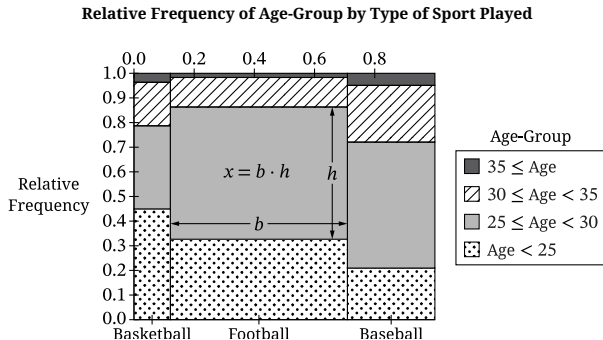
(a)(ii) What is the probability a randomly selected professional athlete is in the age-group of " $25 \leq \text{Age} < 30$ " given they are a football player? Show your work.

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.



Question 5

~ ~ ~
Basketball Football Baseball
Time of Sport

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 ≤ Age < 30		175		1,326		620		2,121		30 ≤ Age < 35		90		287		276		653		35 ≤ 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 = $\text{Age} < 25$, light gray = $25 \leq \text{Age} < 30$, hatched = $30 \leq \text{Age} < 35$, dark gray = $35 \leq \text{Age}$. 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												
$25 \leq \text{Age} < 30$	175	1,326	620	2,121							$30 \leq \text{Age} < 35$	90	287
$35 \leq \text{Age}$	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 " $35 \leq \text{Age}$ " mutually exclusive events? Explain.

(c)(ii) Are the events "Baseball" and " $35 \leq \text{Age}$ " 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 ≤ Age < 30	175	1,326	620	2,121		30 ≤ Age < 35	90	287	276
653		35 ≤ 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 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