

Past-paper walk-through

Introduction to Probability ■ 4.3

eXercise

Questions in this set

- 2026 Q5
- 2025 Q3
- 2024 Q5
- 2019 Q3
- 2017 Q6
- 2014 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

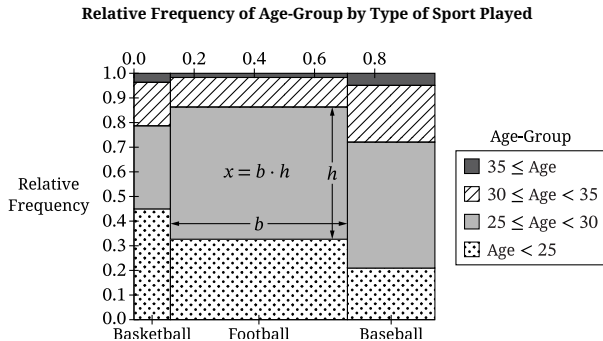
(a)(ii) What is the probability a randomly selected professional athlete is in the age-group of " $25 \leq \text{Age} < 30$ " given they are a football player? Show your work.

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 ≤ 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				

(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												
$25 \leq \text{Age} < 30$	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

(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 3

2025 ■ Q3

Ms. Fey is a manager at a restaurant. To improve the dining experience for her customers, she uses a digital music service to create a playlist of songs that will be played in the restaurant. The playlist contains 1,000 songs and consists of four different types of music in the following quantities: 200 country songs, 400 pop songs, 100 rock songs, and 300 jazz songs. The digital music service will select songs at random from the playlist to be played in the restaurant. Any song can be replayed at any time.

(a)(i) Suppose one song is selected at random to be played. What is the probability that the song is a rock song? Show your work.

(a)(ii) Suppose two songs are selected at random to be played. What is the probability that both songs are rock songs? Show your work.

Question 3

2025 ■ Q3

Ms. Fey is a manager at a restaurant. To improve the dining experience for her customers, she uses a digital music service to create a playlist of songs that will be played in the restaurant. The playlist contains 1,000 songs and consists of four different types of music in the following quantities: 200 country songs, 400 pop songs, 100 rock songs, and 300 jazz songs. The digital music service will select songs at random from the playlist to be played in the restaurant. Any song can be replayed at any time.

(b)(i) In every one-hour period, 20 songs will be played at random and any song can be replayed at any time. Ms. Fey is interested in how many rock songs will be played in a typical one-hour period.

Define the random variable of interest to Ms. Fey, and state how the random variable is distributed.

Question 3

(b)(ii) What is the expected value for the random variable in part B (i)? Show your work.

Question 3

2025 ■ Q3

Ms. Fey is a manager at a restaurant. To improve the dining experience for her customers, she uses a digital music service to create a playlist of songs that will be played in the restaurant.

The playlist contains 1,000 songs and consists of four different types of music in the following quantities: 200 country songs, 400 pop songs, 100 rock songs, and 300 jazz songs. The digital music service will select songs at random from the playlist to be played in the restaurant. Any song can be replayed at any time.

(c)(i) Recall that in every one-hour period, 20 songs will be played at random and any song can be replayed at any time.

Determine the probability that 4 or more rock songs in a particular one-hour period will be played. Show your work.

Question 3

(c)(ii) Suppose 4 rock songs are played during a particular one-hour period. Does this provide strong evidence that the song selection process was not truly random? Justify your answer without performing an inference procedure.

Solution 3

(a)(i)

Mark scheme

- $P(\text{Rock Song}) = 100/1,000 = 0.10$. Full credit needs both: (1) the response calculates the correct probability (0.10), (2) the response shows supporting work for it (the 100/1000 ratio).

Solution 3

(a)(ii) Mark scheme

- $P(\text{Both Rock Songs}) = (0.10)(0.10) = 0.01$, using independence of the two song selections (any song can be replayed / selections are treated as independent). This leaf's credit is bundled with 3(a)(i) in the official scoring (Part A of the SG scores both sub-parts together as one holistic unit), so it carries 0 of this question's marks on its own — but a correct, independence-consistent answer of 0.01 with supporting work is still expected and reinforces credit for part (a). Note: a response that instead samples without replacement, e.g. $(100/1000)(99/999) \approx 0.0099$, is also acceptable if it shows valid supporting work, though it does not reflect the 'each song is replayed randomly' independence framing given in the stem.

Solution 3

(b)(i) Mark scheme

- Let the random variable X represent the number of rock songs (out of the 20 songs played) in one hour. Since any song can be replayed at any time, each of the 20 songs played is independently a rock song with probability $100/1,000 = 0.10$, so X has a binomial distribution with $n = 20$ independent trials and probability of success $p = 0.10$ for each trial, i.e. $X \sim B(20, 0.10)$. Full credit needs at least 3 of 4 components: (1) defines the random variable as the number of rock songs played in one hour (must include both 'number of rock songs' AND 'in one hour'/'out of 20 songs'), (2) describes the distribution as binomial, (3) states $n = 20$ and $p = 0.10$ (may appear in part (b)(i) or (b)(ii)), (4) [see 3(b)(ii)]. Stating songs are 'distributed randomly' is not itself a distribution

Solution 3

statement and should be treated as extraneous; if the response names a non-binomial distribution, this part cannot score essentially correct.

Solution 3

(b)(ii)

Mark scheme

- $E(X) = np = 20(0.10) = 2$ rock songs. This leaf's credit is bundled with 3(b)(i) in the official scoring (Part B of the SG scores both sub-parts together), so it carries 0 of this question's marks on its own — but the correct expected value (2) with the np calculation shown (e.g. $20(0.10) = 2$) is still expected to complete credit for part (b).

Solution 3

(c)(i)

Mark scheme

- $P(X \geq 4) = 1 - P(X \leq 3) = 1 - \left[\binom{20}{0}(0.10)^0(0.90)^{20} + \binom{20}{1}(0.10)^1(0.90)^{19} + \binom{20}{2}(0.10)^2(0.90)^{18} + \binom{20}{3}(0.10)^3(0.90)^{17} \right] = 1 - 0.867 = 0.133$.
 Full credit needs:
 - (1) the correct probability 0.133, (2) supporting work—an explicit binomial probability formula, a labeled bar graph of binomial probabilities, or clearly labeled calculator notations such as $1 - \text{binomcdf}(n = 20, p = 0.10, \text{upperbound} = 3)$ (an unlabeled $\text{binomcdf}(20, 0.10, 3)$ is not sufficient support). An arithmetic/transcription error

Solution 3

(c)(ii) Mark scheme

- No, this does not provide strong evidence that the song-selection process was not truly random. The probability that 4 or more rock songs would be played in a particular one-hour period is 0.133 (from part (c)(i)), which is high enough to be reasonably attributed to random chance alone — it is not small enough to constitute strong evidence against the process being truly random. Full credit needs: (1) indicates there is not a strong reason to believe the selection process was not truly random (i.e., answers 'No'), (2) an explanation that correctly links the (c)(i) probability to that decision (0.133 is not a small/surprising probability). Note: a response that instead argues 'Yes' (there IS strong evidence) can still earn full credit here if it is adequately supported by a correctly-computed alternative

Solution 3

probability or by the standard deviation of the binomial distribution (≈ 1.342) combined with an explanation based on 4 being within two standard deviations of the mean of 2.

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