Past-paper walk-through

Independent Events and Unions of Events ▪ 4.6

eXercise

Questions in this set

- 2026 Q5
- 2025 Q3
- 2019 Q3
- 2017 Q3
- 2017 Q6
- 2016 Q4
- 2014 Q3

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.

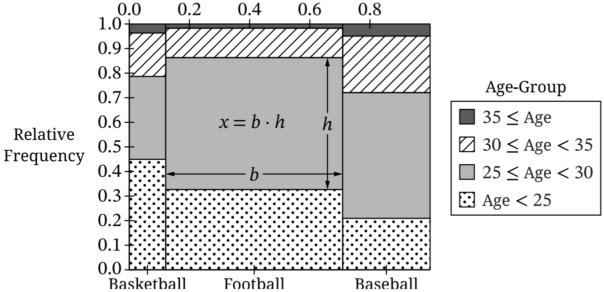
Question 5

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.

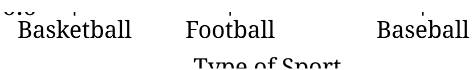
Part B

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
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

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 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(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(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 − $[\binom{20}{0}(0.10)^0(0.90)^{20} + \binom{20}{1}(0.10)^1(0.90)^{19} + \binom{20}{2}(0.10)^2(0.90)^{18} + \binom{20}{3}(0.10)^3(0.90)^{17}] = 1{-}0.867 = 0.133.\mathit{Fullcreditneeds} : (1)\mathit{thecorrectprobability}0.133, (2)\mathit{supportingwork}\text{—}\mathit{anexplicitbinomial} - \mathit{probabilityformula}, \mathit{alabeledbargraphofbinomialprobabilities}, \mathit{orclearly} - \mathit{labeledcalculatornotationsuchas}'1{-}\mathit{binomcdf}(n = 20, p = 0.10, \mathit{upperbound} = 3)'(\mathit{anunlabeled}'\mathit{binomcdf}(20, 0.10, 3)'\mathit{isnotsufficientsupport}).\mathit{Anarithmetic}/\mathit{transcriptionerro}$

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 ($\approx$1.342) combined with an explanation based on 4 being within two standard deviations of the mean of 2.

Question 3

2019 ▪ Q3

A medical researcher surveyed a large group of men and women about whether they take medicine as prescribed. The responses were categorized as never, sometimes, or always. The relative frequency of each category is shown in the table.

	Never	**Sometimes**	**Always**	**Total**
Men	0.0564	0.2016	0.2120	0.4700
Women	0.0636	0.1384	0.3280	0.5300
Total	0.1200	0.3400	0.5400	1.0000

Question 3

(a)(i) One person from those surveyed will be selected at random.
What is the probability that the person selected will be someone whose response is never and who is a woman?
(a)(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?
(a)(iii) What is the probability that the person selected will be someone whose response is never given that the person is a woman?
(b) For the people surveyed, are the events of being a person whose response is never and being a woman independent? Justify your answer.

Question 3

(c) Assume that, in a large population, the probability that a person will always take medicine as prescribed is 0.54. If 5 people are selected at random from the population, what is the probability that at least 4 of the people selected will always take medicine as prescribed? Support your answer.

Question 3

The responses were categorized as never, sometimes, or always. The relative frequency of each category
shown in the table.

	Never	Sometimes	Always	Total
Men	0.0564	0.2016	0.2120	0.4700
Women	0.0636	0.1384	0.3280	0.5300
Total	0.1200	0.3400	0.5400	1.0000

Solution 3

(a)(i)

Mark scheme

- $P(\text{never and woman}) = 0.0636$, read directly as the joint relative frequency from the two-way table.

(a)(ii)

Mark scheme

- $P(\text{never or woman}) = P(\text{never}) + P(\text{woman}) - P(\text{never and woman}) = 0.12 + 0.53 - 0.0636 = 0.5864$. (Equivalently, sum the relative frequencies for all cells that are 'never' or 'woman': $0.0564 + 0.0636 + 0.1384 + 0.3280 = 0.5864$, or $0.0564 + 0.53 = 0.5864$.)

Solution 3

(a)(iii)

Mark scheme

- $P(\text{never} \mid \text{woman}) = \dfrac{P(\text{never and woman})}{P(\text{woman})} = \dfrac{0.0636}{0.53} = 0.12.$

Solution 3

(b)

Mark scheme

- Yes, the event of being a person who responds 'never' is independent of the event of being a woman, because $P(\text{never} \mid \text{woman}) = 0.12 = P(\text{never})$ – the conditional probability equals the unconditional (marginal) probability, using numbers from the table. (An equally valid justification: $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$.) Must state the events are independent AND justify with actual numbers from the table.

Solution 3

(c) Mark scheme

- Let $X =$ the number of people (out of 5 randomly selected) who always take their medicine as prescribed. X has a binomial distribution with $n = 5$ and $p = 0.54$. $P(X \geq 4) = \binom{5}{4}(0.54)^4(0.46)^1 + \binom{5}{5}(0.54)^5(0.46)^0 \approx 0.19557 + 0.04592 \approx 0.24149$. Full credit requires: (1) clearly identifying the binomial distribution with $n = 5, p = 0.54$, (2) correctly identifying the boundary value and direction (at least 4, i.e. $X \geq 4$), and (3) reporting the correct probability ≈ 0.241 (accept values consistent with a correct binomial setup, e.g. 0.24-0.2415). Note: a normal approximation is NOT appropriate here since $np = 5 \times 0.54 = 2.7 < 5$.

Question 3

2017 ▪ Q3

A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

(b) Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K.

For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137 mm?

Question 3

(c) Given that a melon selected at random from the grocery store has a diameter greater than 137 mm, what is the probability that the melon will be from Distributor J?

Solution 3

(a) Mark scheme

- Let X denote the diameter of a randomly selected Distributor J melon. X has an approximately normal distribution with mean 133 mm and standard deviation 5 mm, i.e. $X \sim N(133, 5)$. The z-score for a diameter of 137 mm is $z = (137 - 133)/5 = 4/5 = 0.8$. Therefore $P(X > 137) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$.

Solution 3

- Full credit needs three components: (1) Normality and parameters – indicates use of a normal (or approximately normal) distribution and clearly identifies the correct mean (133) and standard deviation (5); (2) Boundary and direction – uses the correct boundary value ($x = 137$ or $z = 0.8$) and the correct direction (greater than / right tail); (3) Probability – reports the correct normal probability (0.2119) consistent with the setup in (1) and (2).

Solution 3

(b)

Mark scheme

- Define events: $J =$ melon is from Distributor J, $K =$ melon is from Distributor K, $G =$ melon diameter is greater than 137 mm. For a randomly selected melon from the grocery store, by the law of total probability: $P(G) = P(G \mid J)P(J) + P(G \mid K)P(K) = (0.2119)(0.7) + (0.8413)(0.3) = 0.1483 + 0.2524 = 0.4007$. (Equivalently, via a tree diagram: $P(G \text{ and } J) = 0.1483$ and $P(G \text{ and } K) = 0.2524$, which sum to 0.4007.) Full credit needs the probability computed correctly AND work shown using correct numerical values in a correct formula, tree diagram, or table.

Solution 3

(c) Mark scheme

- Using the events defined in part (b), the requested probability is $P(J \mid G) = \dfrac{P(J \text{ and } G)}{P(G)} = \dfrac{P(G \mid J)P(J)}{P(G)} = \dfrac{(0.2119)(0.7)}{0.4007} = \dfrac{0.1483}{0.4007} = 0.3701$. Full credit needs the probability computed correctly AND work shown that illustrates how it was found (the conditional-probability/Bayes formula with numerator and denominator identified, or the equivalent tree-diagram ratio). A correct answer that consistently follows from incorrect part (a)/(b) values (provided all probabilities stay between 0 and 1) still earns credit here.

Question 6

2017 ▪ Q6

Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A	**Treatment**	**Control**
	Man 1	Woman 1
	Man 2	Woman 2

Question 6

Arrangement B	**Treatment**	**Control**
	Man 1 Woman 1	Man 2 Woman 2

Arrangement C	**Treatment**	**Control**
	Man 1 Woman 2	Man 2 Woman 1

Question 6

Arrangement D	Treatment	Control
	Woman 1 Woman 2	Man 1 Man 2

Arrangement E	Treatment	Control
	Man 2 Woman 2	Man 1 Woman 1

Question 6

Arrangement F	Treatment	Control
	Man 2	Man 1
	Woman 1	Woman 2

Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.

Question 6

(a)(i) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.
Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	**A**	**B**	**C**	**D**	**E**	**F**
Probability						

Question 6

(a)(ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

Question 6

Arrangement A	
Treatment	Control
Man 1	Woman 1
Man 2	Woman 2

Arrangement B	
Treatment	Control
Man 1	Man 2
Woman 1	Woman 2

Arrangement C	
Treatment	Control
Man 1	Man 2
Woman 2	Woman 1

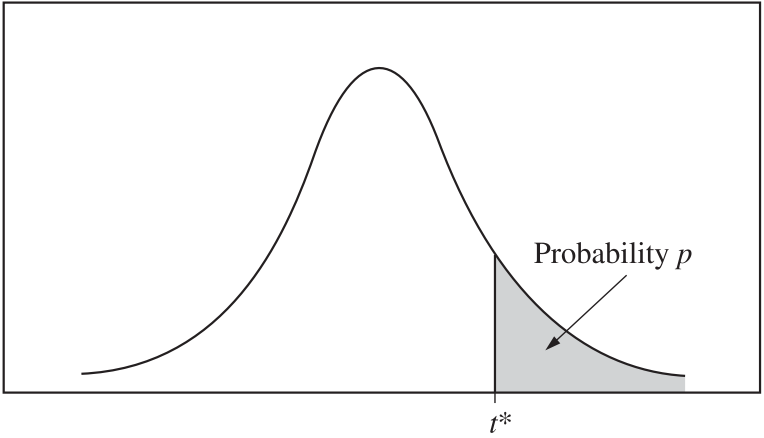
Arrangement D	
Treatment	Control
Woman 1	Man 1
Woman 2	Man 2

Arrangement E	
Treatment	Control
Man 2	Man 1
Woman 2	Woman 1

Arrangement F	
Treatment	Control
Man 2	Man 1
Woman 1	Woman 2

Question 6

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.



Question 6

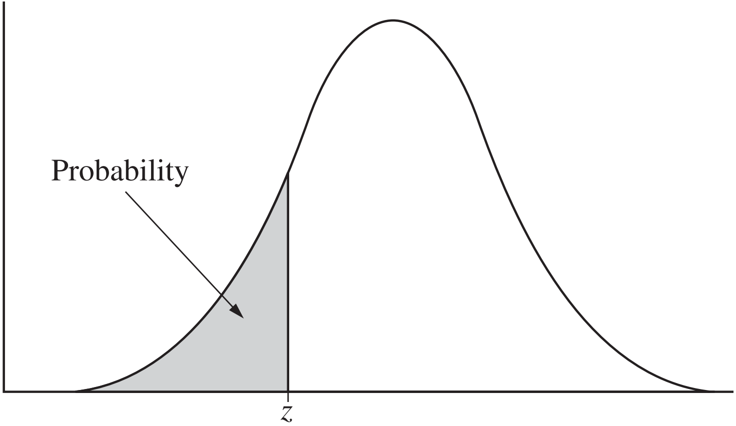


Table entry for z is the probability lying below z.