Week 13

AP Statistics

Homework bundle for this week

本周作业合集

exercises.lol

Contents 目录

4. Probability, Random Variables, and Probability Distributions Starter Quiz · 课前小测

AP Statistics Name: ____________ Score: ______

1. Probability is best described as the long-run...
   - ☐ relative frequency of an outcome
   - ☐ guess about the next trial
   - ☐ number of trials
   - ☐ average of the data
2. In a simulation, 12 of 100 trials gave the outcome of interest. Estimate the probability (as a decimal).

   ____
3. Rolling a fair die, what is P(even)? Give a decimal.

   ____
4. P(A)=0.3, P(B)=0.4, P(A and B)=0.1. Find P(A or B).

   ____
5. P(A and B) = 0.2 and P(B) = 0.5. Find P(A | B).

   ____
6. Two independent events have P(A)=0.5 and P(B)=0.5. Find P(A and B).

   ____
7. The number of heads in 3 coin flips is which kind of random variable?
   - ☐ discrete
   - ☐ continuous
   - ☐ categorical
   - ☐ neither
8. Which is NOT one of the four binomial (BINS) conditions?
   - ☐ the trials must be a random sample of the population
   - ☐ binary outcomes
   - ☐ independent trials
   - ☐ a fixed number of trials

9. After 5 heads in a row on a fair coin, tails is 'due' and more likely on the next flip.

   ☐ True ☐ False
10. The law of large numbers says that as the number of trials grows, the estimated probability...
    - ☐ approaches the true probability
    - ☐ becomes exactly 0.5
    - ☐ gets less reliable
    - ☐ stays fixed forever

1. **relative frequency of an outcome**

   Probability = long-run relative frequency over many repetitions.

2. **0.12**

   Estimate = successes / trials = 12/100 = 0.12.

3. **0.5**

   Even = {2,4,6}, so 3/6 = 0.5.

4. **0.6**

   0.3 + 0.4 − 0.1 = 0.6.

5. **0.4**

   $P(A|B) = P(A \cap B)/P(B) = 0.2/0.5 = 0.4$.

6. **0.25**

   Independent → multiply: $0.5 \times 0.5 = 0.25$.

7. **discrete**

   It takes countable separate values (0,1,2,3) — discrete.

8. **the trials must be a random sample of the population**

   BINS = Binary, Independent, fixed Number, Same p. Random sampling isn't required.

9. **False**

   Each flip is independent — chance has no memory. That's the law-of-averages fallacy.

10. **approaches the true probability**

    More trials → the estimate closes in on the true value.

AP STATISTICS • EXERCISE SHEET

4.1 Introducing Statistics: Random and Non-Random Patterns?

Name: ____________ Class: ________ Date: __________ **Total: 8 marks**

KEY WORDS random 随机 • random process 随机过程 • probability 概率

Objective

Build the skills to answer exam questions on **random and non-random patterns**.

You must be able to:

- explain that a **random process** 随机过程 is unpredictable in the short run but patterned in the long run
- distinguish **random** from **non-random** patterns
- give examples of random processes

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ **Random processes**

A single outcome of a **random process** cannot be predicted, but over **many** repetitions a stable pattern emerges — this long-run regularity is what probability describes.

■ **Random vs non-random**

A **coin flip** is random (each result unpredictable); a fixed daily schedule is **non-random** (predictable).

2 Practice

2.1 State what happens to a random process's behaviour in the long run. [1]

2.2 State the difference between random and non-random patterns. [1]

2.3 Give one example of a random process. [1]

3 Exam-style questions

3.1 A random process is unpredictable in the short run but shows a pattern [1]

- **A** never
- **B** in the long run
- **C** only once
- **D** immediately

3.2 Which of these is a random process? [1]

- **A** the exact sunrise time tomorrow
- **B** a coin flip
- **C** a fixed timetable
- **D** the value of $2 + 2$

3.3 A fair coin is flipped many times.

(a) State whether the result of a single flip is predictable. [1]

(b) State what emerges over many flips. [1]

(c) State the long-run proportion of heads for a fair coin. [1]

Solutions

2.1 it settles into a stable, predictable pattern (relative frequency).

2.2 a random pattern is unpredictable in the short run; a non-random one is predictable.

2.3 a coin flip, a die roll, or drawing a card (any one).

3.1 B.

3.2 B.

3.3 (a) no. (b) a stable pattern — about half heads. (c) 0.5.

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

(a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

(b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

(c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

> Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

AP STATISTICS • EXERCISE SHEET

4.2 Estimating Probabilities Using Simulation

Name: ____________ Class: __________ Date: __________ **Total: 8 marks**

KEY WORDS simulation 模拟 • relative frequency 相对频数 • random 随机 • probability 概率

Objective

Build the skills to answer exam questions on **estimating probabilities using simulation**.

You must be able to:

- use a **simulation 模拟** with random numbers to estimate a probability
- explain how running more trials improves the estimate
- interpret a **relative frequency 相对频数** as an estimate of a probability

1 Worked examples

Study these first. Each one shows the method for a question type used later.

> **Simulation**
>
> A **simulation** uses a **random number generator** to imitate a random process many times. The **relative frequency** of an event (how often it happens ÷ number of trials) estimates its probability.
>
> **More trials, better estimate**
>
> By the long-run pattern of randomness, the estimate gets **closer to the true probability** as the number of simulated trials increases.

2 Practice

2.1 State what a simulation uses to estimate a probability. [1]

2.2 State how the estimate improves. [1]

2.3 State what the relative frequency of an event approximates. [1]

3 Exam-style questions

3.1 A simulation estimates a probability using [1]

- **A** a single trial
- **B** many random trials
- **C** a formula only
- **D** a guess

3.2 As the number of simulated trials increases, the estimate [1]

- **A** gets worse
- **B** gets closer to the true probability
- **C** stays fixed
- **D** becomes 1

3.3 A student simulates rolling a die 1000 times to estimate $P(\text{six})$.

(a) State the tool that provides the randomness. [1]

(b) State what the relative frequency of sixes estimates. [1]

(c) State how to improve the estimate. [1]

Solutions

2.1 many random trials (using a random number generator).

2.2 by running more trials.

2.3 the true probability of the event.

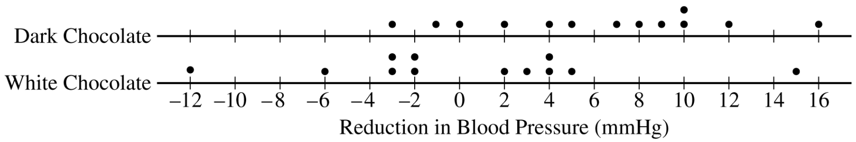
3.1 B.

3.2 B.

3.3 (a) a random number generator. (b) the true probability of rolling a six. (c) run more trials.

. Studies have shown that foods rich in compounds known as flavonoids help lower blood pressure. Researchers conducted a study to investigate whether there was a greater reduction in blood pressure for people who consumed dark chocolate, which contains flavonoids, than people who consumed white chocolate, which does not contain flavonoids. Twenty-five healthy adults agreed to participate in the study and add 3.5 ounces of chocolate to their daily diets. Of the 25 participants, 13 were randomly assigned to the dark chocolate group and the rest were assigned to the white chocolate group. All participants had their blood pressure recorded, in millimeters of mercury (mmHg), before adding chocolate to their daily diets and again 30 days after adding chocolate to their daily diets.

The reduction in blood pressure (before minus after) for each of the participants in the two groups is shown in the dotplots below.



(a) Determine and compare the medians of the reduction in blood pressure for the two groups.

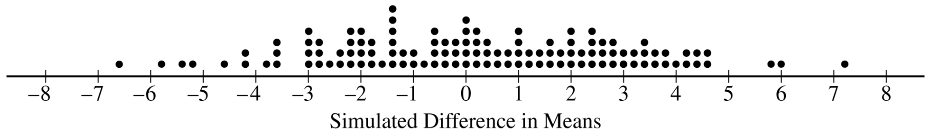
The researchers found the mean reduction in blood pressure for those who consumed dark chocolate is $\bar{x}_{dark} = 6.08$ mmHg and the mean reduction in blood pressure for those who consumed white chocolate is $\bar{x}_{white} = 0.42$ mmHg.

(b) One researcher indicated that because the difference in sample means of 5.66 mmHg is greater than 0 there is convincing statistical evidence to conclude that the population mean blood pressure reduction for those who consume dark chocolate is greater than for those who consume white chocolate. Why might the researcher's conclusion, based only on the difference in sample means of 5.66 mmHg, not necessarily be true?

GO ON TO THE NEXT PAGE.

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

A simulation was conducted to investigate whether there is a greater reduction of blood pressure for those who consume dark chocolate than for those who consume white chocolate. The simulation was conducted under the assumption that no difference exists. The results of 120 trials of the simulation are shown in the following dotplot.



(c) Use the results of the simulation to determine whether the results from the 25 participants in the study provide convincing statistical evidence, at a 5 percent level of significance, that adding dark chocolate to a daily diet will result in a greater reduction in blood pressure, on average, than adding white chocolate to a daily diet. Justify your answer.

GO ON TO THE NEXT PAGE.

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

   (c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

   > Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

   Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

AP STATISTICS • EXERCISE SHEET

4.3 Introduction to Probability

Name: ____________ Class: __________ Date: ___________ **Total: 9 marks**

KEY WORDS complement 补 • sample space 样本空间 • probability 概率 • random 随机

Objective

Build the skills to answer exam questions on **the basics of probability**.

You must be able to:

- identify the **sample space** 样本空间 of a random process
- calculate $P(E) = \dfrac{\text{outcomes in } E}{\text{total outcomes}}$ for equally likely outcomes
- use that a probability is between 0 and 1
- find the **complement**, $P(E') = 1 - P(E)$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Basic probability

The **sample space** is the set of all possible outcomes. If all are equally likely:

$$P(E) = \frac{\text{number of outcomes in } E}{\text{total number of outcomes}}.$$

Range and complement

Every probability satisfies $0 \leq P(E) \leq 1$. The **complement** (not E) has $P(E') = 1 - P(E)$. A probability is the **long-run relative frequency** of the event.

Example

A fair die: $P(\text{even}) = \dfrac{3}{6} = 0.5$.

2 Practice

2.1 State the range of any probability. [1]

2.2 A fair die is rolled. Find $P(\text{rolling a }4)$. [1]

2.3 If $P(\text{rain}) = 0.3$, find $P(\text{no rain})$. [2]

3 Exam-style questions

3.1 The probability of the complement of E is [1]

- **A** $P(E)$
- **B** $1 - P(E)$
- **C** $1 + P(E)$
- **D** 0

3.2 A probability must be between [1]

- **A** -1 and 1
- **B** 0 and 1
- **C** 0 and 100
- **D** any two values

3.3 A bag has 3 red and 7 blue balls, drawn at random.

(a) Find $P(\text{red})$. [2]

(b) Find $P(\text{not red})$. [1]

Solutions

2.1 between 0 and 1 (inclusive).

2.2 $\frac{1}{6} \approx 0.167$.

2.3 $P(\text{no rain}) = 1 - 0.3 = 0.7$.

3.1 **B**.

3.2 **B**.

3.3 (a) $P(\text{red}) = \frac{3}{10} = 0.3$. (b) $1 - 0.3 = 0.7$.

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
$25 \leq$ 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

Part A

Consider the two-way table.

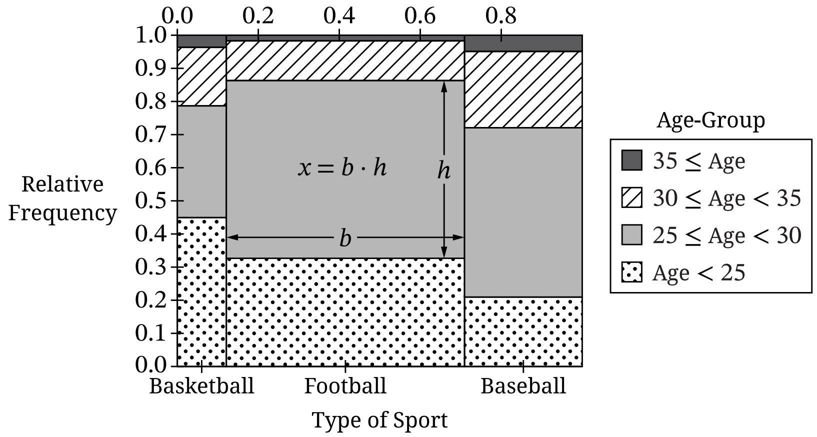
i. What is the probability a randomly selected professional athlete is a football player? Show your work.

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.

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



i. 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)?

ii. What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

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

ii. Are the events "Baseball" and "$35 \leq$ 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.

3. 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.

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

B. 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.

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

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

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

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

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.

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.

3. 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

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(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.

(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.

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. 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

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

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.

(a) 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.

(i) 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						

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

The six arrangements are repeated below.

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

(b) For the chip method, two chips are marked "treatment" and two chips are marked "control." Each person selects one chip at random without replacement.

(i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

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

(c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

STOP

END OF EXAM

Table entry for z is the probability lying below z.

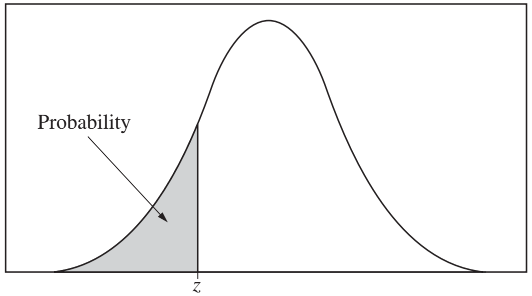


Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table entry for z is the probability lying below z.

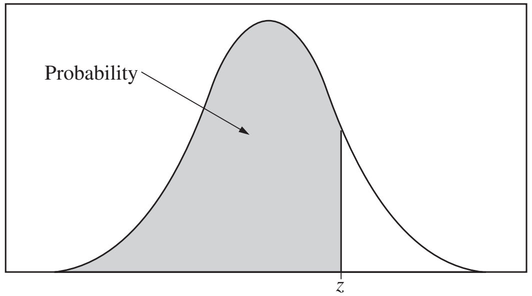
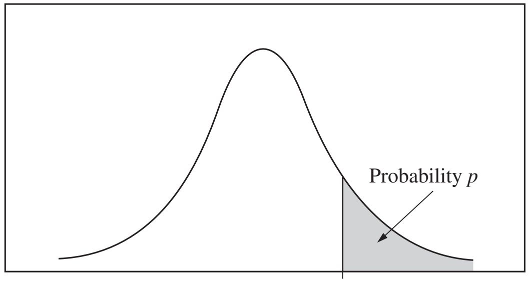


Table A *(Continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

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

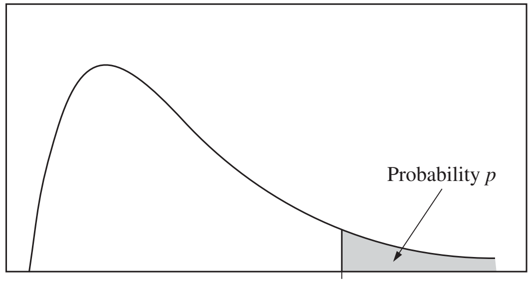


t^*

Table B ***t* distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Table entry for p is the point (χ^2) with probability p lying above it.



(χ^2)

Table C **χ^2 critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

   (c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

   Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

AP STATISTICS • EXERCISE SHEET

4.4 Mutually Exclusive Events

Name: ______ Class: ______ Date: ______ **Total: 8 marks**

KEY WORDS mutually exclusive 互斥 • joint probability 联合概率 • probability 概率 • complement 补

Objective

Build the skills to answer exam questions on **mutually exclusive events**.

You must be able to:

- describe the **joint probability** 联合概率 $P(A \cap B)$
- explain why two events are or are not **mutually exclusive** 互斥
- use that mutually exclusive (disjoint) events have $P(A \cap B) = 0$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Joint probability

$P(A \cap B)$ is the probability that **both** A and B occur (the intersection).

Mutually exclusive events

Two events are **mutually exclusive** (disjoint) if they **cannot happen at the same time**, so

$$P(A \cap B) = 0.$$

For example, on one die roll, "roll a 2" and "roll a 5" are mutually exclusive.

2 Practice

2.1 State what $P(A \cap B)$ means. [1]

2.2 Define mutually exclusive events. [1]

2.3 State $P(A \cap B)$ for two mutually exclusive events. [1]

3 Exam-style questions

3.1 Two mutually exclusive events have $P(A \cap B) =$ [1]

- **A** 1
- **B** 0
- **C** 0.5
- **D** $P(A) \cdot P(B)$

3.2 $P(A \cap B)$ is called the ________ probability. [1]

- **A** conditional
- **B** joint
- **C** marginal
- **D** complement

3.3 On one die roll, event $A =$ "roll a 3" and event $B =$ "roll a 6".

(a) State whether A and B are mutually exclusive. [1]

(b) State $P(A \cap B)$. [1]

(c) Find $P(A)$. [1]

Solutions

2.1 the probability that both A and B occur.

2.2 events that cannot occur at the same time.

2.3 0.

3.1 **B**.

3.2 **B**.

3.3 (a) yes. (b) 0. (c) $\frac{1}{6}$.

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
$25 \leq$ 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

Part A

Consider the two-way table.

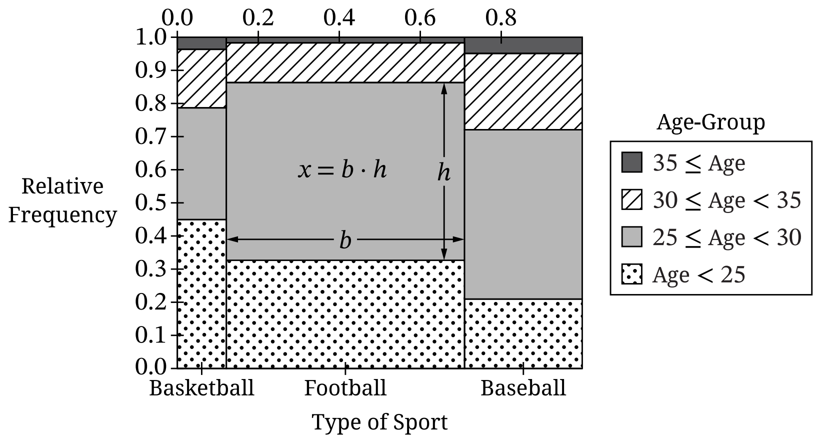
i. What is the probability a randomly selected professional athlete is a football player? Show your work.

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.

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



i. 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)?

ii. What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

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

ii. Are the events "Baseball" and "$35 \leq$ 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.

AP STATISTICS • EXERCISE SHEET

4.5 Conditional Probability

Name: ____________ Class: ________ Date: __________ **Total: 9 marks**

KEY WORDS multiplication rule 乘法法则 • conditional probability 条件概率 • probability 概率

Objective

Build the skills to answer exam questions on **conditional probability**.

You must be able to:

- calculate a **conditional probability** 条件概率, $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
- apply the **multiplication rule** 乘法法则, $P(A \cap B) = P(A) \cdot P(B \mid A)$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

is the probability of A **given** that B has occurred.

Multiplication rule

$$P(A \cap B) = P(A) \cdot P(B \mid A).$$

Example

If $P(A \cap B) = 0.2$ and $P(B) = 0.5$, then $P(A \mid B) = \dfrac{0.2}{0.5} = 0.4$.

2 Practice

2.1 Write the formula for $P(A \mid B)$. [1]

2.2 If $P(A \cap B) = 0.12$ and $P(B) = 0.4$, find $P(A \mid B)$. [2]

2.3 State the multiplication rule. [1]

3 Exam-style questions

3.1 $P(A \mid B)$ equals [1]

- **A** $\dfrac{P(A \cap B)}{P(B)}$
- **B** $P(A) + P(B)$
- **C** $P(A) \cdot P(B)$
- **D** $\dfrac{P(B)}{P(A)}$

3.2 The multiplication rule gives $P(A \cap B) =$ [1]

- **A** $P(A) + P(B \mid A)$
- **B** $P(A) \cdot P(B \mid A)$
- **C** $\dfrac{P(A)}{P(B)}$
- **D** $P(A) - P(B)$

3.3 In a class, $P(\text{likes maths and science}) = 0.3$ and $P(\text{science}) = 0.6$.

(a) Find $P(\text{likes maths} \mid \text{science})$. [2]

(b) Write the general formula for $P(A \mid B)$. [1]

Solutions

2.1 $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.

2.2 $P(A \mid B) = \dfrac{0.12}{0.4} = 0.3$.

2.3 $P(A \cap B) = P(A) \cdot P(B \mid A)$.

3.1 **A**.

3.2 **B**.

3.3 (a) $\dfrac{0.3}{0.6} = 0.5$. (b) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.

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
$25 \leq$ 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

Part A

Consider the two-way table.

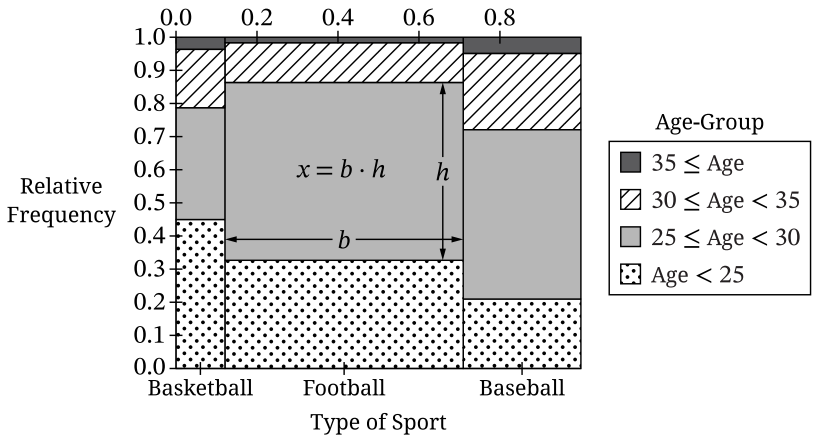
i. What is the probability a randomly selected professional athlete is a football player? Show your work.

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.

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



i. 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)?

ii. What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

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

ii. Are the events "Baseball" and "$35 \leq$ 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.

Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n} \qquad s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\hat{y} = a + bx \qquad \bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) \qquad b = r\frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum(x_i - \mu_X)^2 P(x_i)}$
If X has a **binomial** distribution with parameters n and p, then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a **geometric** distribution with parameter p, then: $P(X = x) = (1-p)^{x-1}p$ where $x = 1, 2, 3, \ldots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ where $\hat{p}_c = \dfrac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\overline{X}$	$\mu_{\overline{X}} = \mu$	$\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$	$s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$
For two populations: $\overline{X}_1 - \overline{X}_2$	$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$	$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$, where $\sigma_x = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n}}$	$s_b = \dfrac{s}{s_x\sqrt{n-1}}$, where $s = \sqrt{\dfrac{\Sigma(y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n-1}}$

andard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the ndard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

3. 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

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(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.

(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.

3. Approximately 3.5 percent of all children born in a certain region are from multiple births (that is, twins, triplets, etc.). Of the children born in the region who are from multiple births, 22 percent are left-handed. Of the children born in the region who are from single births, 11 percent are left-handed.

(a) What is the probability that a randomly selected child born in the region is left-handed?

(b) What is the probability that a randomly selected child born in the region is a child from a multiple birth, given that the child selected is left-handed?

(c) A random sample of 20 children born in the region will be selected. What is the probability that the sample will have at least 3 children who are left-handed?

AP Statistics: **2017**

3. 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?

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.

(b) 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?

(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?

STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. 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 Man 2	Woman 1 Woman 2

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

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

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

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

Arrangement F	
Treatment	Control
Man 2 Woman 1	Man 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.

(a) 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.

(i) 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						

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

The six arrangements are repeated below.

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

(b) For the chip method, two chips are marked "treatment" and two chips are marked "control." Each person selects one chip at random without replacement.

(i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

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

(c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

STOP

END OF EXAM

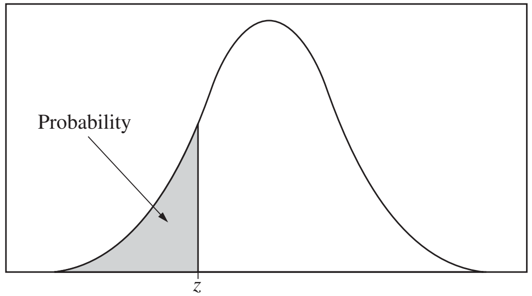


Table entry for z is the probability lying below z.

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

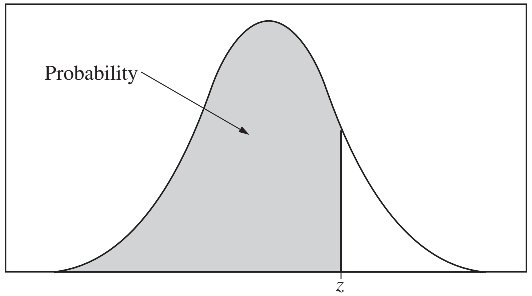


Table entry for z is the probability lying below z.

Table A ***(Continued)***

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

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

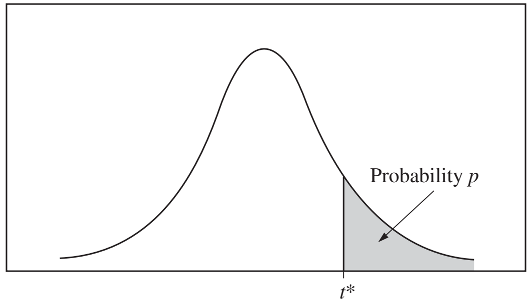


Table B ***t*** **distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Table entry for p is the point (χ^2) with probability p lying above it.

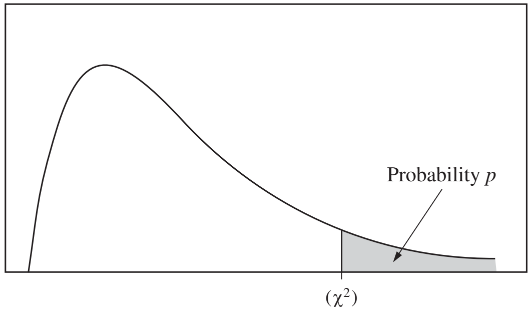


Table C χ^2 **critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

3. A shopping mall has three automated teller machines (ATMs). Because the machines receive heavy use, they sometimes stop working and need to be repaired. Let the random variable X represent the number of ATMs that are working when the mall opens on a randomly selected day. The table shows the probability distribution of X.

Number of ATMs working when the mall opens	0	1	2	3
Probability	0.15	0.21	0.40	0.24

(a) What is the probability that at least one ATM is working when the mall opens?

(b) What is the expected value of the number of ATMs that are working when the mall opens?

(c) What is the probability that all three ATMs are working when the mall opens, given that at least one ATM is working?

(d) Given that at least one ATM is working when the mall opens, would the expected value of the number of ATMs that are working be less than, equal to, or greater than the expected value from part (b) ? Explain.

AP STATISTICS • EXERCISE SHEET

4.6 Independent Events and Unions of Events

Name: ____________ Class: ________ Date: __________ **Total: 9 marks**

KEY WORDS independent 独立 • addition rule 加法法则 • probability 概率

Objective

Build the skills to answer exam questions on **independent events and unions of events**.

You must be able to:

- describe **independence** 独立: $P(A \mid B) = P(A)$, so $P(A \cap B) = P(A) \cdot P(B)$
- apply the **addition rule** 加法法则, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Independence

A and B are **independent** if knowing that one occurred does not change the probability of the other: $P(A \mid B) = P(A)$. Then

$$P(A \cap B) = P(A) \cdot P(B).$$

■ The addition rule

For the **union** (A or B or both):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

2 Practice

2.1 State the condition for A and B to be independent. [1]

2.2 For independent events with $P(A) = 0.5$ and $P(B) = 0.4$, find $P(A \cap B)$. [2]

2.3 Write the addition rule for $P(A \cup B)$. [1]

3 Exam-style questions

3.1 For independent events, $P(A \cap B) =$ [1]

- **A** $P(A) + P(B)$
- **B** $P(A) \cdot P(B)$
- **C** 0
- **D** $P(A \mid B)$

3.2 The addition rule is $P(A \cup B) =$ [1]

- **A** $P(A) + P(B)$
- **B** $P(A) + P(B) - P(A \cap B)$
- **C** $P(A) \cdot P(B)$
- **D** $P(A) - P(B)$

3.3 $P(A) = 0.6$, $P(B) = 0.3$, and $P(A \cap B) = 0.2$.

(a) Find $P(A \cup B)$. [2]

(b) State whether A and B are independent (check $P(A \cap B) = P(A) \cdot P(B)$). [1]

Solutions

2.1 knowing that one has occurred does not change the probability of the other.

2.2 $P(A \cap B) = 0.5 \times 0.4 = 0.2$.

2.3 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

3.1 **B**.

3.2 **B**.

3.3 (a) $0.6 + 0.3 - 0.2 = 0.7$. (b) no — $P(A) \cdot P(B) = 0.18 \neq 0.2$.

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 ≤ 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

Part A

Consider the two-way table.

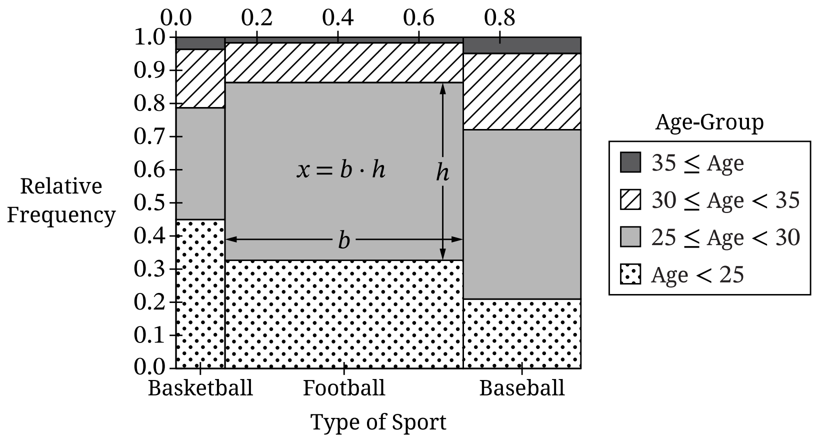
i. What is the probability a randomly selected professional athlete is a football player? Show your work.

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



i. 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)?

ii. What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

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

ii. Are the events "Baseball" and "$35 \leq$ 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.

3. 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.

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

B. 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.

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

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

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

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

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.

3. 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

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(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.

(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.

3. 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?

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.

(b) 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?

(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?

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. 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 Man 2	Woman 1 Woman 2

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

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

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

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

Arrangement F	
Treatment	Control
Man 2 Woman 1	Man 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.

(a) 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.

(i) 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						

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

The six arrangements are repeated below.

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

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

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

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

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

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

(b) For the chip method, two chips are marked “treatment” and two chips are marked “control.” Each person selects one chip at random without replacement.

(i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

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

(c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

STOP

END OF EXAM

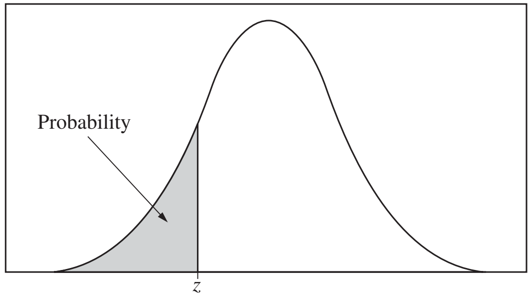


Table entry for z is the probability lying below z.

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

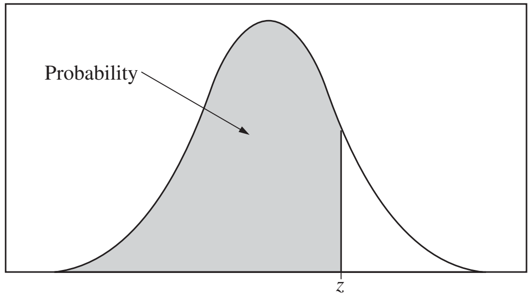


Table entry for z is the probability lying below z.

Table A (*Continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

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

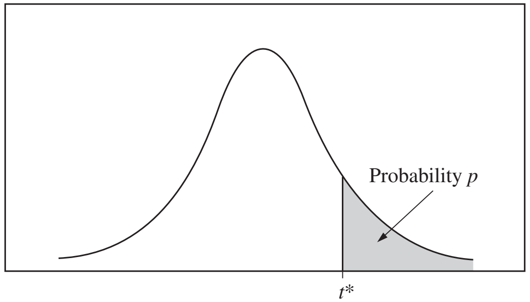


Table B ***t*** **distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Name: AP Statistics: **2017**

Table entry for p is the point (χ^2) with probability p lying above it.

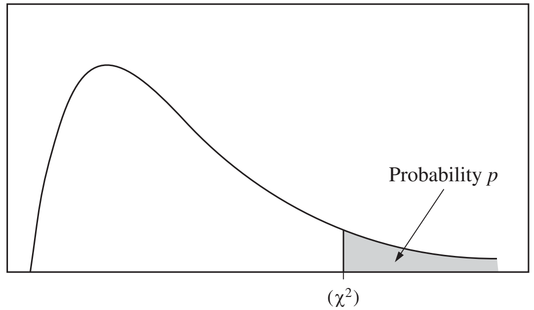


Table C χ^2 **critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

4. A company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.

   A company engineer develops a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

   Step 1: One super igniter is selected at random and used in a rocket.

   Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

   Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.

   (a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?

   (b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first or the thirty-second super igniter tested if the failure rate of the super igniters is 15 percent?

   (c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

   (c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

   Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

---

3. Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. The daily number of absences at High School A in the state is approximately normally distributed with mean of 120 students and standard deviation of 10.5 students.

   (a) If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year. Approximately what is the probability that High School A will lose some state funding?

   (b) The principals' association in the state suggests that instead of choosing one day at random, the state should choose 3 days at random. With the suggested plan, High School A would lose some of its state funding in the subsequent year if the mean number of students absent for the 3 days is greater than 140. Would High School A be more likely, less likely, or equally likely to lose funding using the suggested plan compared to the plan described in part (a)? Justify your choice.

   (c) A typical school week consists of the days Monday, Tuesday, Wednesday, Thursday, and Friday. The principal at High School A believes that the number of absences tends to be greater on Mondays and Fridays, and there is concern that the school will lose state funding if the attendance count occurs on a Monday or Friday. If one school day is chosen at random from each of 3 typical school weeks, what is the probability that none of the 3 days chosen is a Tuesday, Wednesday, or Thursday?

4.7 Introduction to Random Variables and Probability Distributions

Name: ____________ Class: ________ Date: __________ **Total: 9 marks**

KEY WORDS random variable 随机变量 • probability distribution 概率分布 • discrete random variable 离散随机变量 • random 随机 • probability 概率

Objective

Build the skills to answer exam questions on **random variables and probability distributions**.

You must be able to:

- describe the values of a **random variable** 随机变量 as numerical outcomes
- explain that a **discrete random variable** 离散随机变量 has countable values whose probabilities sum to 1
- represent a **probability distribution** 概率分布 as a table, graph, or function

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Random variables

The values of a **random variable** are the numerical outcomes of a random process. A **discrete** random variable can take only a **countable** number of values, each with a probability.

■ The key rule

The probabilities of all possible values must **sum to** 1. A **probability distribution** lists the values and their probabilities (as a table, graph, or function).

2 Practice

2.1 State what the probabilities in a discrete distribution must sum to. [1]

2.2 A random variable X has $P(0) = 0.2$, $P(1) = 0.5$, and $P(2) = ?$. Find the missing probability. [2]

2.3 Define a discrete random variable. [1]

3 Exam-style questions

3.1 The probabilities of a discrete random variable sum to [1]

- **A** 0
- **B** 1
- **C** 100
- **D** the mean

3.2 A discrete random variable takes [1]

- **A** any real value
- **B** a countable number of values
- **C** only 0 or 1
- **D** only negative values

3.3 X has $P(1) = 0.3$, $P(2) = 0.4$, and $P(3) = ?$.

(a) Find $P(3)$. [2]

(b) State what all the probabilities must sum to. [1]

Solutions

2.1 1.

2.2 $P(2) = 1 - 0.2 - 0.5 = 0.3$.

2.3 a variable that can take only a countable number of numerical values, each with a probability.

3.1 **B**.

3.2 **B**.

3.3 (a) $P(3) = 1 - 0.3 - 0.4 = 0.3$. (b) 1.

5. According to a 2017 national survey in Country B, the mean number of bedrooms in newly built houses was 2.9. Rodney, a researcher, believes the mean number of bedrooms in newly built houses in the country was different in 2024 than it was in 2017. To investigate his belief, he took a large random sample of newly built houses in Country B in 2024 and recorded the number of bedrooms in each house. The distribution of the number of bedrooms for the sampled houses is summarized in the table.

Distribution of the Number of Bedrooms for the Houses Sampled in 2024

Number of Bedrooms	1	2	3	4	5	6
Proportion of Houses	0.12	0.22	0.28	0.22	0.14	0.02

A.

i. A house from the sample will be selected at random. What is the probability that the house had fewer than 3 bedrooms? Show your work.

ii. What is the mean number of bedrooms for the sample of newly built houses in 2024? Show your work.

B. Rodney will use a one-sample t-test for a population mean to test his belief.

i. In the context of Rodney's investigation, state the hypotheses for the test.

ii. Explain, in context, what a Type I error would be for Rodney's hypothesis test.

C. A different researcher, Keisha, suggests using a confidence interval to investigate whether the mean number of bedrooms in newly built houses in 2024 in Country B was different from 2.9.

Assume the conditions for inference have been met. Using Rodney's data, Keisha calculated a one-sample 97 percent confidence interval to estimate the population mean as $(3.01, 3.19)$. Based on the confidence interval, what conclusion can be made for Rodney's hypothesis test in part B at $\alpha = 0.03$? Justify your answer.

Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\hat{y} = a + bx$$

$$\bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b = r\frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a **binomial** distribution with parameters n and p, then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a **geometric** distribution with parameter p, then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \ldots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

.

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x\sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x\sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

andard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the ıdard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

3. A shopping mall has three automated teller machines (ATMs). Because the machines receive heavy use, they sometimes stop working and need to be repaired. Let the random variable X represent the number of ATMs that are working when the mall opens on a randomly selected day. The table shows the probability distribution of X.

Number of ATMs working when the mall opens	0	1	2	3
Probability	0.15	0.21	0.40	0.24

(a) What is the probability that at least one ATM is working when the mall opens?

(b) What is the expected value of the number of ATMs that are working when the mall opens?

(c) What is the probability that all three ATMs are working when the mall opens, given that at least one ATM is working?

(d) Given that at least one ATM is working when the mall opens, would the expected value of the number of ATMs that are working be less than, equal to, or greater than the expected value from part (b) ? Explain.

AP STATISTICS • EXERCISE SHEET

4.8 Mean and Standard Deviation of Random Variables

Name: ______ Class: ______ Date: ______ **Total: 9 marks**

KEY WORDS mean (expected value) 期望值 • random 随机 • probability 概率 • random variable 随机变量

Objective

Build the skills to answer exam questions on **the mean and standard deviation of random variables**.

You must be able to:

- calculate the **mean** (expected value) 期望值 $\mu_X = \sum x_i \cdot P(x_i)$
- calculate the **standard deviation** $\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$
- interpret these parameters in context

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Mean (expected value)

$$\mu_X = \sum x_i \cdot P(x_i)$$

— a probability-weighted average of the values.

■ Standard deviation

$$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}.$$

■ Example

X: 0 (prob 0.5), 1 (prob 0.3), 2 (prob 0.2):

$$\mu_X = 0(0.5) + 1(0.3) + 2(0.2) = 0.7.$$

2 Practice

2.1 Write the formula for the mean of a discrete random variable. [1]

2.2 X takes 1 (prob 0.5) and 3 (prob 0.5). Find the mean. [2]

2.3 State what the mean of a random variable is also called. [1]

3 Exam-style questions

3.1 The expected value of X is [1]

- **A** $\sum x_i$
- **B** $\sum x_i \cdot P(x_i)$
- **C** $\sum P(x_i)$
- **D** the largest value of x

3.2 The mean of a random variable is also called its [1]

- **A** median
- **B** expected value
- **C** mode
- **D** variance

3.3 X takes 2 (prob 0.4) and 5 (prob 0.6).

(a) Find the mean. [2]

(b) State what the probabilities sum to. [1]

Solutions

2.1 $\mu_X = \sum x_i \cdot P(x_i)$.

2.2 $\mu_X = 1(0.5) + 3(0.5) = 2$.

2.3 the expected value.

3.1 B.

3.2 B.

3.3 (a) $\mu_X = 2(0.4) + 5(0.6) = 0.8 + 3 = 3.8$. (b) 1.

5. According to a 2017 national survey in Country B, the mean number of bedrooms in newly built houses was 2.9. Rodney, a researcher, believes the mean number of bedrooms in newly built houses in the country was different in 2024 than it was in 2017. To investigate his belief, he took a large random sample of newly built houses in Country B in 2024 and recorded the number of bedrooms in each house. The distribution of the number of bedrooms for the sampled houses is summarized in the table.

Distribution of the Number of Bedrooms for the Houses Sampled in 2024

Number of Bedrooms	1	2	3	4	5	6
Proportion of Houses	0.12	0.22	0.28	0.22	0.14	0.02

A.

i. A house from the sample will be selected at random. What is the probability that the house had fewer than 3 bedrooms? Show your work.

ii. What is the mean number of bedrooms for the sample of newly built houses in 2024? Show your work.

B. Rodney will use a one-sample *t*-test for a population mean to test his belief.

i. In the context of Rodney's investigation, state the hypotheses for the test.

ii. Explain, in context, what a Type I error would be for Rodney's hypothesis test.

C. A different researcher, Keisha, suggests using a confidence interval to investigate whether the mean number of bedrooms in newly built houses in 2024 in Country B was different from 2.9.

Assume the conditions for inference have been met. Using Rodney's data, Keisha calculated a one-sample 97 percent confidence interval to estimate the population mean as $(3.01, 3.19)$. Based on the confidence interval, what conclusion can be made for Rodney's hypothesis test in part B at $\alpha = 0.03$? Justify your answer.

4. In an online game, players move through a virtual world collecting geodes, a type of hollow rock. When broken open, these geodes contain crystals of different colors that are useful in the game. A red crystal is the most useful crystal in the game. The color of the crystal in each geode is independent and the probability that a geode contains a red crystal is 0.08.

(a) Sarah, a player, will collect and open geodes until a red crystal is found.

(i) Calculate the mean of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(ii) Calculate the standard deviation of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(b) Another player, Conrad, decides to play the game and will stop opening geodes after finding a red crystal or when 4 geodes have been opened, whichever comes first. Let $Y =$ the number of geodes Conrad will open. The table shows the partially completed probability distribution for the random variable Y.

Number of geodes Conrad will open, y	1	2	3	4
Probability, $P(Y = y)$	0.08	0.0736		

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 4** on this page.

(i) Calculate $P(Y = 3)$. Show your work.

(ii) Calculate $P(Y = 4)$. Show your work.

(c) Consider the table and your results from part (b).

(i) Calculate the mean of the distribution of the number of geodes Conrad will open. Show your work.

(ii) Interpret the mean of the distribution of the number of geodes Conrad will open, which was calculated in part (c-i).

GO ON TO THE NEXT PAGE.

Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n} \qquad s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\hat{y} = a + bx \qquad \bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) \qquad b = r\frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a **binomial** distribution with parameters n and p, then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a **geometric** distribution with parameter p, then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \ldots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

andard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the ndard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

(b) Consider two rules for identifying outliers, method A and method B. Let method A represent the $1.5 \times \text{IQR}$ rule, and let method B represent the 2 standard deviations rule.

(i) Using method A, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(ii) The mean length of stay for the sample is 7.42 days with a standard deviation of 2.37 days. Using method B, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(c) Explain why method A might identify more data points as potential outliers than method B for a distribution that is strongly skewed to the right.

GO ON TO THE NEXT PAGE.

2. Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(a) One aspect of the study is to record the number of miles each subject walks per day. The researchers are deciding whether to have subjects wear an activity tracker to record the data or to have subjects keep a daily journal of the miles they walk each day. Describe what bias could be introduced by keeping the daily journal instead of wearing the activity tracker.

GO ON TO THE NEXT PAGE.

During the course of the study, the subjects will have their cholesterol levels measured each month by a doctor. The researchers will perform a significance test at the end of the study to determine whether the average cholesterol level for subjects who walk fewer miles each day is greater than for those who walk more miles each day.

(b) Selecting a random sample creates a reasonable representative sample of the target population. Explain the benefit of using a representative sample from the population.

(c) Suppose the researchers conduct the test and find a statistically significant result. Would it be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population? Explain your reasoning.

GO ON TO THE NEXT PAGE.

3. To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year.

(i) Define the random variable of interest and state how the random variable is distributed.

(ii) Determine the probability that a particular employee receives at least one gift card in a 52-week year. Show your work.

GO ON TO THE NEXT PAGE.

(b) Calculate and interpret the expected value for the number of gift cards a particular employee will receive in a 52-week year. Show your work.

(c) Suppose that Agatha, an employee at the company, never receives a gift card for an entire 52-week year. Based on her experience, does Agatha have a strong argument that the selection process was not truly random? Explain your answer.

GO ON TO THE NEXT PAGE.

5. A company that manufactures smartphones developed a new battery that has a longer life span than that of a traditional battery. From the date of purchase of a smartphone, the distribution of the life span of the new battery is approximately normal with mean 30 months and standard deviation 8 months. For the price of $50, the company offers a two-year warranty on the new battery for customers who purchase a smartphone. The warranty guarantees that the smartphone will be replaced at no cost to the customer if the battery no longer works within 24 months from the date of purchase.

   (a) In how many months from the date of purchase is it expected that 25 percent of the batteries will no longer work? Justify your answer.

   (b) Suppose one customer who purchases the warranty is selected at random. What is the probability that the customer selected will require a replacement within 24 months from the date of purchase because the battery no longer works?

   (c) The company has a gain of $50 for each customer who purchases a warranty but does not require a replacement. The company has a loss (negative gain) of $150 for each customer who purchases a warranty and does require a replacement. What is the expected value of the gain for the company for each warranty purchased?

3. A shopping mall has three automated teller machines (ATMs). Because the machines receive heavy use, they sometimes stop working and need to be repaired. Let the random variable X represent the number of ATMs that are working when the mall opens on a randomly selected day. The table shows the probability distribution of X.

Number of ATMs working when the mall opens	0	1	2	3
Probability	0.15	0.21	0.40	0.24

(a) What is the probability that at least one ATM is working when the mall opens?

(b) What is the expected value of the number of ATMs that are working when the mall opens?

(c) What is the probability that all three ATMs are working when the mall opens, given that at least one ATM is working?

(d) Given that at least one ATM is working when the mall opens, would the expected value of the number of ATMs that are working be less than, equal to, or greater than the expected value from part (b) ? Explain.

AP STATISTICS • EXERCISE SHEET

4.9 Combining Random Variables

Name: ____________ Class: __________ Date: __________ **Total: 9 marks**

KEY WORDS random variable 随机变量 • random 随机 • independent 独立

Objective

Build the skills to answer exam questions on **combining random variables**.

You must be able to:

- find the mean of $aX + bY$ as $a\mu_X + b\mu_Y$
- find the variance of $aX + bY$ for **independent** X, Y as $a^2\sigma_X^2 + b^2\sigma_Y^2$
- describe a **linear transformation** $Y = a + bX$: $\mu_Y = a + b\mu_X$, $\sigma_Y = |b|\sigma_X$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ **Means add**

$$\mu_{aX+bY} = a\mu_X + b\mu_Y \quad \text{(always).}$$

■ **Variances add (if independent)**

$$\sigma^2_{aX+bY} = a^2\sigma_X^2 + b^2\sigma_Y^2 \quad \text{(independent).}$$

Note that **variances**, not standard deviations, add.

■ **Linear transformation**

For $Y = a + bX$: $\mu_Y = a + b\mu_X$ and $\sigma_Y = |b|\sigma_X$ (adding a constant does not change the spread).

2 Practice

2.1 State the mean of $X + Y$ in terms of μ_X and μ_Y. [1]

2.2 If $\mu_X = 5$ and $\mu_Y = 3$, find the mean of $X + Y$. [1]

2.3 For $Y = 2X$ with $\sigma_X = 4$, find σ_Y. [2]

3 Exam-style questions

3.1 The mean of $X + Y$ is [1]

- **A** $\mu_X \cdot \mu_Y$
- **B** $\mu_X + \mu_Y$
- **C** $\mu_X - \mu_Y$
- **D** μ_X / μ_Y

3.2 For **independent** X and Y, the variance of $X + Y$ is [1]

- **A** $\sigma_X + \sigma_Y$
- **B** $\sigma_X^2 + \sigma_Y^2$
- **C** $\sigma_X \cdot \sigma_Y$
- **D** $\sigma_X^2 - \sigma_Y^2$

3.3 X has $\mu_X = 10$ and $\sigma_X = 2$. Let $Y = 3X$.

(a) Find μ_Y. [1]

(b) Find σ_Y. [2]

Solutions

2.1 $\mu_X + \mu_Y$.

2.2 $5 + 3 = 8$.

2.3 $\sigma_Y = |2|\sigma_X = 2 \times 4 = 8$.

3.1 **B**.

3.2 **B**.

3.3 (a) $\mu_Y = 3 \times 10 = 30$. (b) $\sigma_Y = |3| \times 2 = 6$.

Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n} \qquad s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\hat{y} = a + bx \qquad \bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) \qquad b = r\frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum(x_i - \mu_X)^2 P(x_i)}$
If X has a **binomial** distribution with parameters n and p, then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a **geometric** distribution with parameter p, then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \ldots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum\dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x\sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x\sqrt{n-1}}$, where $s = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$

andard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the ndard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

AP STATISTICS • EXERCISE SHEET

4.10 Introduction to the Binomial Distribution

Name: ____________ Class: _________ Date: __________ **Total: 11 marks**

KEY WORDS binomial 二项 • random 随机 • probability 概率 • independent 独立 • random variable 随机变量

Objective

Build the skills to answer exam questions on **the binomial distribution**.

You must be able to:

- recognise a **binomial** 二项 setting: n independent trials, each a success or failure, with fixed p
- calculate $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ The binomial setting

A binomial random variable X counts the number of **successes** in n trials that are:

- **fixed** in number (n),
- **two-outcome** (success/failure),
- **independent**, with
- a **constant** probability of success p.

■ The binomial formula

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

For $n = 3$, $p = 0.5$, $x = 2$: $\binom{3}{2}(0.5)^2(0.5)^1 = 3 \times 0.125 = 0.375$.

2 Practice

2.1 State the four conditions of a binomial setting. [2]

2.2 Write the binomial probability formula. [1]

2.3 For $n = 4$, $p = 0.5$, find $P(X = 4)$. [2]

3 Exam-style questions

3.1 A binomial random variable counts the number of [1]

- **A** trials
- **B** successes in n trials
- **C** failures only
- **D** outcomes

3.2 In a binomial setting, the probability of success p is [1]

- **A** changing each trial
- **B** constant across trials
- **C** zero
- **D** exactly 1

3.3 A fair coin is flipped 3 times; X = number of heads.

(a) State n and p. [1]

(b) Find $P(X = 0)$. [2]

(c) Name the distribution. [1]

Solutions

2.1 fixed number of trials; two outcomes; independent trials; constant p.

2.2 $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$.

2.3 $P(X = 4) = \binom{4}{4}(0.5)^4 = (0.5)^4 = 0.0625$.

3.1 **B**.

3.2 **B**.

3.3 (a) $n = 3$, $p = 0.5$. (b) $P(X = 0) = (0.5)^3 = 0.125$. (c) binomial.

3. A certain sports team has its team song performed by a different musician at the beginning of each of its games. The time it takes for the team song to be performed by different musicians can be modeled using a normal distribution with mean $\mu = 109$ seconds and standard deviation $\sigma = 16$ seconds. All performances of the team song are independent.

Part A

What is the probability that a randomly selected performance of the team song will take longer than 120 seconds? Show your work.

Part B

Ten performances of the team song will be randomly selected. Let the random variable X represent the number of games at which the performance of the team song takes longer than 120 seconds. Find $P(X \geq 3)$. Show your work.

Part C

Ben will attend all of the games for this sports team. Let the random variable Y represent the number of games Ben will attend until a performance of the team song takes longer than 120 seconds.

i. Calculate the mean of Y. Show your work.

ii. Calculate the standard deviation of Y. Show your work.

Part D

Interpret the standard deviation calculated in part C (ii) in context.

3. 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.

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

B. 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.

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

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

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

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

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.

. A machine at a manufacturing company is programmed to fill shampoo bottles such that the amount of shampoo in each bottle is normally distributed with mean 0.60 liter and standard deviation 0.04 liter. Let the random variable A represent the amount of shampoo, in liters, that is inserted into a bottle by the filling machine.

(a) A bottle is considered to be underfilled if it has less than 0.50 liter of shampoo. Determine the probability that a randomly selected bottle of shampoo will be underfilled. Show your work.

After the bottles are filled, they are placed in boxes of 10 bottles per box. After the bottles are placed in the boxes, several boxes are placed in a crate for shipping to a beauty supply warehouse. The manufacturing company's contract with the beauty supply warehouse states that one box will be randomly selected from a crate. If 2 or more bottles in the selected box are underfilled, the entire crate will be rejected and sent back to the manufacturing company.

(b) The beauty supply warehouse manager is interested in the probability that a crate shipped to the warehouse will be rejected. Assume that the amounts of shampoo in the bottles are independent of each other.

(i) Define the random variable of interest for the warehouse manager and state how the random variable is distributed.

(ii) Determine the probability that a crate will be rejected by the warehouse manager. Show your work.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 3** on this page.

To reduce the number of crates rejected by the beauty supply warehouse manager, the manufacturing company is considering adjusting the programming of the filling machine so that the amount of shampoo in each bottle is normally distributed with mean 0.56 liter and standard deviation 0.03 liter.

(c) Would you recommend that the manufacturing company use the original programming of the filling machine or the adjusted programming of the filling machine? Provide a statistical justification for your choice.

GO ON TO THE NEXT PAGE.

(b) Consider two rules for identifying outliers, method A and method B. Let method A represent the $1.5 \times \text{IQR}$ rule, and let method B represent the 2 standard deviations rule.

(i) Using method A, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(ii) The mean length of stay for the sample is 7.42 days with a standard deviation of 2.37 days. Using method B, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(c) Explain why method A might identify more data points as potential outliers than method B for a distribution that is strongly skewed to the right.

GO ON TO THE NEXT PAGE.

2. Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(a) One aspect of the study is to record the number of miles each subject walks per day. The researchers are deciding whether to have subjects wear an activity tracker to record the data or to have subjects keep a daily journal of the miles they walk each day. Describe what bias could be introduced by keeping the daily journal instead of wearing the activity tracker.

GO ON TO THE NEXT PAGE.

During the course of the study, the subjects will have their cholesterol levels measured each month by a doctor. The researchers will perform a significance test at the end of the study to determine whether the average cholesterol level for subjects who walk fewer miles each day is greater than for those who walk more miles each day.

(b) Selecting a random sample creates a reasonable representative sample of the target population. Explain the benefit of using a representative sample from the population.

(c) Suppose the researchers conduct the test and find a statistically significant result. Would it be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population? Explain your reasoning.

GO ON TO THE NEXT PAGE.

3. To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year.

(i) Define the random variable of interest and state how the random variable is distributed.

(ii) Determine the probability that a particular employee receives at least one gift card in a 52-week year. Show your work.

GO ON TO THE NEXT PAGE.

(b) Calculate and interpret the expected value for the number of gift cards a particular employee will receive in a 52-week year. Show your work.

(c) Suppose that Agatha, an employee at the company, never receives a gift card for an entire 52-week year. Based on her experience, does Agatha have a strong argument that the selection process was not truly random? Explain your answer.

GO ON TO THE NEXT PAGE.

3. 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

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(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.

(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.

2. An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

Do you regularly recycle plastic bottles?

Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as $(0.584, 0.816)$.

(a) How many students were in the sample selected by the environmental science teacher?

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

(c) The statistics teacher at the high school was concerned about the potential bias in the survey. To obtain a potentially less biased estimate of the proportion, the statistics teacher used an alternate method for collecting student responses. A random sample of 300 students was selected, and each student was given the following instructions on how to respond to the question.

- In private, flip a fair coin.
- If heads, you must respond no, regardless of whether you regularly recycle.
- If tails, please truthfully respond yes or no.

(i) What is the expected number of students from the sample of 300 who would be required to respond no because the coin flip resulted in heads?

(ii) The results of the sample showed that 213 of the 300 selected students responded no. Based on the results of the sample, give a point estimate for the proportion of all students at the high school who would respond yes to the question.

3. Approximately 3.5 percent of all children born in a certain region are from multiple births (that is, twins, triplets, etc.). Of the children born in the region who are from multiple births, 22 percent are left-handed. Of the children born in the region who are from single births, 11 percent are left-handed.

(a) What is the probability that a randomly selected child born in the region is left-handed?

(b) What is the probability that a randomly selected child born in the region is a child from a multiple birth, given that the child selected is left-handed?

(c) A random sample of 20 children born in the region will be selected. What is the probability that the sample will have at least 3 children who are left-handed?

AP STATISTICS • EXERCISE SHEET

4.11 Parameters for a Binomial Distribution

Name: ____________ Class: ________ Date: __________ **Total: 11 marks**

KEY WORDS random 随机 • probability 概率 • random variable 随机变量 • binomial 二项

Objective

Build the skills to answer exam questions on **the parameters of a binomial distribution**.

You must be able to:

- calculate the **mean** of a binomial, $\mu_X = np$
- calculate the **standard deviation**, $\sigma_X = \sqrt{np(1-p)}$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Binomial parameters

For a binomial random variable with n trials and success probability p:

$$\mu_X = np, \qquad \sigma_X = \sqrt{np(1-p)}.$$

■ Example

$n = 100$, $p = 0.2$: $\mu_X = 100(0.2) = 20$ and $\sigma_X = \sqrt{100(0.2)(0.8)} = \sqrt{16} = 4$.

2 Practice

2.1 Write the formula for the mean of a binomial distribution. [1]

2.2 For $n = 50$, $p = 0.4$, find the mean. [2]

2.3 For $n = 100$, $p = 0.5$, find the standard deviation. [2]

3 Exam-style questions

3.1 The mean of a binomial distribution is [1]

- **A** n
- **B** np
- **C** p
- **D** $\sqrt{np}$

3.2 The standard deviation of a binomial distribution is [1]

- **A** np
- **B** $\sqrt{np(1-p)}$
- **C** $n(1-p)$
- **D** p

3.3 A binomial distribution has $n = 200$ and $p = 0.3$.

(a) Find the mean. [2]

(b) Find the standard deviation. [2]

Solutions

2.1 $\mu_X = np$.

2.2 $\mu_X = 50 \times 0.4 = 20$.

2.3 $\sigma_X = \sqrt{100(0.5)(0.5)} = \sqrt{25} = 5$.

3.1 **B**.

3.2 **B**.

3.3 (a) $\mu_X = 200 \times 0.3 = 60$. (b) $\sigma_X = \sqrt{200(0.3)(0.7)} = \sqrt{42} \approx 6.5$.

3. A certain sports team has its team song performed by a different musician at the beginning of each of its games. The time it takes for the team song to be performed by different musicians can be modeled using a normal distribution with mean $\mu = 109$ seconds and standard deviation $\sigma = 16$ seconds. All performances of the team song are independent.

Part A

What is the probability that a randomly selected performance of the team song will take longer than 120 seconds? Show your work.

Part B

Ten performances of the team song will be randomly selected. Let the random variable X represent the number of games at which the performance of the team song takes longer than 120 seconds. Find $P(X \geq 3)$. Show your work.

Part C

Ben will attend all of the games for this sports team. Let the random variable Y represent the number of games Ben will attend until a performance of the team song takes longer than 120 seconds.

i. Calculate the mean of Y. Show your work.

ii. Calculate the standard deviation of Y. Show your work.

Part D

Interpret the standard deviation calculated in part C (ii) in context.

3. 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.

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

B. 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.

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

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

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

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

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.

. A machine at a manufacturing company is programmed to fill shampoo bottles such that the amount of shampoo in each bottle is normally distributed with mean 0.60 liter and standard deviation 0.04 liter. Let the random variable A represent the amount of shampoo, in liters, that is inserted into a bottle by the filling machine.

(a) A bottle is considered to be underfilled if it has less than 0.50 liter of shampoo. Determine the probability that a randomly selected bottle of shampoo will be underfilled. Show your work.

After the bottles are filled, they are placed in boxes of 10 bottles per box. After the bottles are placed in the boxes, several boxes are placed in a crate for shipping to a beauty supply warehouse. The manufacturing company's contract with the beauty supply warehouse states that one box will be randomly selected from a crate. If 2 or more bottles in the selected box are underfilled, the entire crate will be rejected and sent back to the manufacturing company.

(b) The beauty supply warehouse manager is interested in the probability that a crate shipped to the warehouse will be rejected. Assume that the amounts of shampoo in the bottles are independent of each other.

(i) Define the random variable of interest for the warehouse manager and state how the random variable is distributed.

(ii) Determine the probability that a crate will be rejected by the warehouse manager. Show your work.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 3** on this page.

To reduce the number of crates rejected by the beauty supply warehouse manager, the manufacturing company is considering adjusting the programming of the filling machine so that the amount of shampoo in each bottle is normally distributed with mean 0.56 liter and standard deviation 0.03 liter.

(c) Would you recommend that the manufacturing company use the original programming of the filling machine or the adjusted programming of the filling machine? Provide a statistical justification for your choice.

GO ON TO THE NEXT PAGE.

(b) Consider two rules for identifying outliers, method A and method B. Let method A represent the $1.5 \times \text{IQR}$ rule, and let method B represent the 2 standard deviations rule.

(i) Using method A, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(ii) The mean length of stay for the sample is 7.42 days with a standard deviation of 2.37 days. Using method B, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(c) Explain why method A might identify more data points as potential outliers than method B for a distribution that is strongly skewed to the right.

GO ON TO THE NEXT PAGE.

Name: AP Statistics: **2021**

2. Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(a) One aspect of the study is to record the number of miles each subject walks per day. The researchers are deciding whether to have subjects wear an activity tracker to record the data or to have subjects keep a daily journal of the miles they walk each day. Describe what bias could be introduced by keeping the daily journal instead of wearing the activity tracker.

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.

During the course of the study, the subjects will have their cholesterol levels measured each month by a doctor. The researchers will perform a significance test at the end of the study to determine whether the average cholesterol level for subjects who walk fewer miles each day is greater than for those who walk more miles each day.

(b) Selecting a random sample creates a reasonable representative sample of the target population. Explain the benefit of using a representative sample from the population.

(c) Suppose the researchers conduct the test and find a statistically significant result. Would it be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population? Explain your reasoning.

GO ON TO THE NEXT PAGE.

3. To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year.

(i) Define the random variable of interest and state how the random variable is distributed.

(ii) Determine the probability that a particular employee receives at least one gift card in a 52-week year. Show your work.

GO ON TO THE NEXT PAGE.

(b) Calculate and interpret the expected value for the number of gift cards a particular employee will receive in a 52-week year. Show your work.

(c) Suppose that Agatha, an employee at the company, never receives a gift card for an entire 52-week year. Based on her experience, does Agatha have a strong argument that the selection process was not truly random? Explain your answer.

GO ON TO THE NEXT PAGE.

3. 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

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(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.

(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.

3. Approximately 3.5 percent of all children born in a certain region are from multiple births (that is, twins, triplets, etc.). Of the children born in the region who are from multiple births, 22 percent are left-handed. Of the children born in the region who are from single births, 11 percent are left-handed.
   (a) What is the probability that a randomly selected child born in the region is left-handed?
   (b) What is the probability that a randomly selected child born in the region is a child from a multiple birth, given that the child selected is left-handed?
   (c) A random sample of 20 children born in the region will be selected. What is the probability that the sample will have at least 3 children who are left-handed?

AP Statistics: **2016**

4. A company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.

   A company engineer develops a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

   Step 1: One super igniter is selected at random and used in a rocket.

   Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

   Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.
   (a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?
   (b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first or the thirty-second super igniter tested if the failure rate of the super igniters is 15 percent?
   (c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.

AP STATISTICS • EXERCISE SHEET

4.12 The Geometric Distribution

Name: ______ Class: ______ Date: ______ **Total: 9 marks**

KEY WORDS geometric 几何 • independent 独立 • probability 概率 • random 随机 • random variable 随机变量

Objective

Build the skills to answer exam questions on **the geometric distribution**.

You must be able to:

- describe a **geometric 几何** random variable as the number of trials until the first success
- use $P(\text{first success on trial } x) = (1-p)^{x-1}p$
- find the mean, $\dfrac{1}{p}$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ The geometric setting

Like a binomial, but instead of a fixed number of trials, a **geometric** random variable counts how many **trials until the first success** (independent trials, constant p).

■ Probability and mean

$$P(X = x) = (1-p)^{x-1}p, \qquad \mu_X = \frac{1}{p}.$$

With $p = 0.2$, the expected number of trials is $\dfrac{1}{0.2} = 5$.

2 Practice

2.1 State what a geometric random variable counts. [1]

2.2 For $p = 0.25$, find the mean number of trials to the first success. [2]

2.3 Write P(first success on trial 1) in terms of p. [1]

3 Exam-style questions

3.1 A geometric random variable counts the [1]

- **A** number of successes in n trials
- **B** number of trials until the first success
- **C** total number of trials
- **D** number of failures only

3.2 The mean of a geometric distribution is [1]

- **A** p
- **B** $\frac{1}{p}$
- **C** np
- **D** $\sqrt{p}$

3.3 A game gives $P(\text{win}) = 0.1$ on each try.

(a) Find the expected number of tries until the first win. [2]

(b) Name the distribution. [1]

Solutions

2.1 the number of trials until the first success.

2.2 $\mu_X = \frac{1}{0.25} = 4$.

2.3 $P(X = 1) = p$.

3.1 B.

3.2 B.

3.3 (a) $\frac{1}{0.1} = 10$. (b) geometric.

3. A certain sports team has its team song performed by a different musician at the beginning of each of its games. The time it takes for the team song to be performed by different musicians can be modeled using a normal distribution with mean $\mu = 109$ seconds and standard deviation $\sigma = 16$ seconds. All performances of the team song are independent.

Part A

What is the probability that a randomly selected performance of the team song will take longer than 120 seconds? Show your work.

Part B

Ten performances of the team song will be randomly selected. Let the random variable X represent the number of games at which the performance of the team song takes longer than 120 seconds. Find $P(X \geq 3)$. Show your work.

Part C

Ben will attend all of the games for this sports team. Let the random variable Y represent the number of games Ben will attend until a performance of the team song takes longer than 120 seconds.

i. Calculate the mean of Y. Show your work.

ii. Calculate the standard deviation of Y. Show your work.

Part D

Interpret the standard deviation calculated in part C (ii) in context.

4. In an online game, players move through a virtual world collecting geodes, a type of hollow rock. When broken open, these geodes contain crystals of different colors that are useful in the game. A red crystal is the most useful crystal in the game. The color of the crystal in each geode is independent and the probability that a geode contains a red crystal is 0.08.

(a) Sarah, a player, will collect and open geodes until a red crystal is found.

(i) Calculate the mean of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(ii) Calculate the standard deviation of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(b) Another player, Conrad, decides to play the game and will stop opening geodes after finding a red crystal or when 4 geodes have been opened, whichever comes first. Let $Y =$ the number of geodes Conrad will open. The table shows the partially completed probability distribution for the random variable Y.

Number of geodes Conrad will open, y	1	2	3	4
Probability, $P(Y = y)$	0.08	0.0736		

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 4** on this page.

(i) Calculate $P(Y = 3)$. Show your work.

(ii) Calculate $P(Y = 4)$. Show your work.

(c) Consider the table and your results from part (b).

(i) Calculate the mean of the distribution of the number of geodes Conrad will open. Show your work.

(ii) Interpret the mean of the distribution of the number of geodes Conrad will open, which was calculated in part (c-i).

GO ON TO THE NEXT PAGE.

4. A company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.

   A company engineer develops a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

   Step 1: One super igniter is selected at random and used in a rocket.

   Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

   Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.

   (a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?

   (b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first or the thirty-second super igniter tested if the failure rate of the super igniters is 15 percent?

   (c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.