

PRACTICE WORKSHEET + FULL ANSWER KEY

Data, probability and interpretation

Secondary statistics. Work through the sections in order, or choose the one that matches this week's topic.

QUESTIONS	SECTIONS	ANSWER KEY	WORKING SPACE
104	10	At the back	Ruled

Name _____ Class _____ Date _____

Before you start

Start with the question and how the data were collected. Summaries describe a sample; a claim about a larger population also depends on sampling and variability.

WORKED EXAMPLE

Find the mean and median of 2, 3, 4, 6, 10.

- 1 The total is 25 and there are 5 values, so mean = 5.
- 2 The ordered middle value is 4, so median = 4.
- 3 The high value of 10 pulls the mean above the median.

Use this method as your model. Set out your own working the same way.

Common mistakes to avoid

- Treating correlation as causation
- Ignoring whether a sample represents the population

There are 104 questions across 10 sections. Write in the ruled space, then mark your work against the answer key at the back. Topics covered: Data summaries, Probability, Evaluating conclusions.

SECTION A Summarising data

Order the data first, then choose the summary that answers the question asked.

1

Find the mean of 6, 8, 10, 12.

2

Find the mean and the median of 2, 3, 4, 6, 10.

3

Find the median of 4, 9, 2, 7, 5, 8.

4

Find the mode of 3, 7, 7, 9, 12, 7.

5

Find the range of 4, 7, 7, 10, 12.

6

A class records travel times of 20, 25, 25, 30 and 40 minutes. Find the mean and the median.

7

Five numbers have a mean of 12. Find their total.

8

Four numbers have a mean of 15. A fifth value of 25 is added. Find the new mean.

9

Explain when the median describes a data set better than the mean.

10

Find the mean of 3, 5, 5, 6, 21 and comment on the effect of the 21.

11

Find the mean of 12, 15, 15, 18, 20, 22 to 1 decimal place.

SECTION B Spread and displays

Choose a display that suits the type of data, and label every axis with its units.

12

Find the quartiles and the median of 2, 4, 5, 7, 8, 11, 13.

13

Find the interquartile range of that data set.

14

Explain what a box plot shows that a bar chart does not.

15

Would you use a bar chart or a histogram for shoe sizes recorded as whole numbers? Explain.

16

Describe the appearance of a scatter plot showing strong negative correlation.

17

A stem-and-leaf plot has stem 3 with leaves 1, 4 and 4. List the values.

18

Find the mean from this frequency table: 1 occurs twice, 2 occurs three times, 3 occurs five times.

19

State one advantage of a pie chart.

20

Two data sets have the same mean but different ranges. What does that tell you?

21

Identify the outlier in 5, 6, 6, 7, 30 and describe its effect on the mean.

22

Explain why the vertical axis of a bar chart should usually start at zero.

SECTION C Probability

Probabilities lie between 0 and 1. List the possible outcomes whenever you are unsure.

23

A fair die is rolled. Find $P(\text{a number greater than 4})$.

24

A bag holds 3 red and 2 blue counters. Find $P(\text{red})$.

25

Two fair coins are tossed. Find $P(\text{exactly one head})$.

26

A fair die is rolled. Find $P(\text{an even number})$.

27

A spinner has 8 equal sectors numbered 1 to 8. Find $P(\text{a multiple of 3})$.

28

$P(\text{rain}) = 0.35$. Find $P(\text{no rain})$.

29

A bag holds 4 green, 5 yellow and 1 white counter. Find $P(\text{not yellow})$.

30

A card is drawn from a standard 52-card pack. Find $P(\text{a heart})$.

31

Two fair dice are rolled. Find $P(\text{a total of 7})$.

32

An event has probability 0.15. How many times is it expected in 200 trials?

33

Explain the difference between theoretical and experimental probability.

SECTION D Combined events and two-way tables

Draw a table or a tree diagram, and decide whether the events are independent.

- 34 Two fair coins are tossed. List all four outcomes.

- 35 A coin is tossed and a die is rolled. How many outcomes are there?

- 36 Find $P(\text{two heads})$ when a fair coin is tossed twice.

- 37 A bag holds 3 red and 5 blue counters. Two are drawn with replacement. Find $P(\text{both red})$.

- 38 The same bag, but drawn without replacement. Find $P(\text{both red})$.

- 39 In a class of 30, 18 study French and 12 study Spanish, with 5 studying both. How many study neither?

- 40 For that class, find $P(\text{a student chosen at random studies French only})$.

- 41 A table shows 40 girls (25 walk) and 60 boys (30 walk). Find $P(\text{a walker chosen at random is a girl})$.

- 42 For that table, find $P(\text{a student chosen at random is a boy who does not walk})$.

- 43 Explain what it means for two events to be independent.

- 44 Give an example of two events that are not independent.

SECTION E Interpreting and evaluating claims

A summary describes a sample. A claim about a population also depends on how that sample was chosen.

- 45 A survey of a school football team finds high interest in sport. Can it represent all students?

- 46 Explain why correlation does not establish causation.

- 47 Give one advantage of a larger sample.

- 48 Describe how you would take a simple random sample of 30 students from 600.

49

Explain one weakness of a voluntary online survey.

50

A report says sales are up 50% but gives no starting figure. What else do you need?

51

A shop finds a mean spend of \$42 and a median spend of \$27. What does that suggest?

52

Two groups have means of 61 and 63 with a large spread. Can you conclude one is better?

53

Name the problem with: Do you agree that our excellent library should open longer?

54

State two things to check before trusting a statistic in an advertisement.

SECTION F Standard deviation and variance

Find the mean first, then work with the distance of each value from it.

55

Find the mean of 2, 4, 4, 4, 5, 5, 7, 9.

56

For that data set, find the variance.

57

For that data set, find the standard deviation.

58

State what a large standard deviation tells you.

59

Two classes share a mean but have standard deviations 2 and 8. Compare them.

60

Every value in a data set increases by 5. What happens to the mean and the standard deviation?

61

Every value is doubled. What happens to the mean and the standard deviation?

62

Explain why deviations from the mean are squared.

63

Find the range and the mean of 10, 12, 14, 16, 18.

64

State one advantage of the standard deviation over the range.

SECTION G Sampling and study design

Ask who could have been left out of the sample, and what that would do to the result.

65 Define a population and a sample.

66 Describe stratified sampling.

67 Describe systematic sampling.

68 A school has 400 girls and 600 boys. In a stratified sample of 50, how many boys?

69 Name one source of bias in a street survey run at 10 am on a weekday.

70 Explain why a census is not always practical.

71 State one advantage of a random sample over a convenience sample.

72 Explain the difference between an experiment and an observational study.

73 Explain why a control group must be treated identically apart from the variable studied.

74 A survey achieves a 12% response rate. Explain the concern.

SECTION H Correlation and lines of best fit

Always look at the scatter plot as well as the number.

75 Describe a correlation coefficient of 0.9.

76 Describe a correlation coefficient of -0.15.

77 State the range of possible values for a correlation coefficient.

78 A line of best fit is $y = 2.5x + 4$. Predict y when $x = 6$.

79 Explain why predicting y at $x = 100$ from that line may be unsafe.

80

Ice cream sales and drowning both rise in summer. Give the likely explanation.

81

Explain what a residual is.

82

Describe what data with no correlation looks like on a scatter plot.

83

State one thing a scatter plot shows that a correlation coefficient cannot.

84

Two variables have $r = 0$. Can they be related? Explain.

SECTION I Conditional probability and tree diagrams

Draw the tree and write each probability on its branch before multiplying.

85

A bag has 4 red and 6 blue counters. Two are drawn without replacement. Find $P(\text{both blue})$.

86

For the same bag and draws, find $P(\text{one of each colour})$.

87

For that pair of draws, find $P(\text{at least one red})$.

88

A test is 95% accurate. Explain why a positive result is not a 95% chance of having the condition.

89

Two events are mutually exclusive with $P(A) = 0.3$ and $P(B) = 0.5$. Find $P(A \text{ or } B)$.

90

For independent events with $P(A) = 0.4$ and $P(B) = 0.25$, find $P(A \text{ and } B)$.

91

Write the formula for the conditional probability of A given B.

92

In a class, $P(\text{sport}) = 0.6$ and $P(\text{sport and music}) = 0.24$. Find $P(\text{music given sport})$.

93

A 4-sector spinner is spun twice. Find $P(\text{the same result twice})$.

94

Explain why a tree diagram helps with dependent events.

SECTION J Mixed review and reading reports

Say what the number describes, and what it does not.

- 95 Find the median of 3, 8, 12, 15, 21, 25.

- 96 Find the mode and range of 4, 4, 7, 9, 9, 9, 12.

- 97 Estimate the mean from grouped data: 10 values with midpoint 5 and 30 with midpoint 15.

- 98 A box plot has minimum 4, Q1 8, median 12, Q3 20, maximum 30. Find the interquartile range.

- 99 For that box plot, describe the skew.

- 100 A cumulative frequency graph reaches 80 at the top. At what frequency do you read the median?

- 101 A headline says a drug doubles your risk. What else do you need to know?

- 102 Explain why a mean can mislead for house prices.

- 103 A poll of 1,000 reports 52% with a 3% margin of error. Can you conclude a majority?

- 104 State two questions to ask about any statistic in a news report.

Answer key

Work each section before you check it. These are model answers: an equivalent method, or a different well-supported interpretation, can be just as correct.

SECTION A - SUMMARISING DATA

1. 9.
2. Mean 5 and median 4.
3. 6, the mean of the middle values 5 and 7.
4. 7.
5. 8.
6. Mean 28 minutes and median 25 minutes.
7. 60.
8. 17, because the total is 85 across 5 values.
9. When extreme values are present, because the median is not pulled towards them.
10. The mean is 8, larger than four of the five values, because 21 is an outlier.
11. 17.0.

SECTION B - SPREAD AND DISPLAYS

12. $Q1 = 4$, median = 7 and $Q3 = 11$.
13. 7.
14. It shows the median, quartiles and range, so it displays spread and skew.
15. A bar chart, because the data are discrete categories.
16. The points lie close to a line sloping downwards from left to right.
17. 31, 34 and 34.
18. 2.3, from a total of 23 across 10 values.
19. It shows each category as a share of the whole.
20. They have a similar centre but different spread.
21. 30 is the outlier; it raises the mean to 10.8, above every other value.
22. A truncated axis exaggerates the differences between the bars.

SECTION C - PROBABILITY

23. $2/6$, which simplifies to $1/3$.
24. $3/5$.
25. $1/2$, from HT and TH out of four outcomes.
26. $1/2$.
27. $2/8$, which simplifies to $1/4$.
28. 0.65.
29. $5/10$, which simplifies to $1/2$.
30. $13/52$, which simplifies to $1/4$.
31. $6/36$, which simplifies to $1/6$.
32. 30 times.

33. Theoretical probability comes from equally likely outcomes; experimental probability comes from observed results and varies between samples.

SECTION D - COMBINED EVENTS AND TWO-WAY TABLES

34. HH, HT, TH and TT.
35. 12.
36. $1/4$.
37. $9/64$.
38. $3/28$, from $(3/8) \times (2/7)$.
39. 5 students, because $18 + 12 - 5 = 25$ study at least one.
40. $13/30$.
41. $25/55$, which simplifies to $5/11$.
42. $30/100$, which simplifies to $3/10$.
43. The outcome of one event does not change the probability of the other.
44. Drawing two cards without replacement; the first draw changes what remains.

SECTION E - INTERPRETING AND EVALUATING CLAIMS

45. Not reliably; the sample overrepresents students already interested in sport.
46. Two variables can move together because of a third factor or by chance; a controlled comparison is needed for a causal claim.
47. It usually reduces the effect of unusual individuals and gives a more stable estimate.
48. Number the students 1 to 600 and use random numbers to select 30 without repeats.
49. Only motivated people respond, so the sample may not represent the population.
50. The original sales figure and the time period, so the size of the change is clear.
51. A few large purchases pull the mean above the median, so the distribution is skewed.
52. Not from the means alone; the spread and sample size mean the difference may not be meaningful.
53. It is a leading question; the wording encourages agreement.
54. Who collected the data and how, and whether the sample and the comparison are appropriate.

SECTION F - STANDARD DEVIATION AND VARIANCE

55. 5.
56. 4.
57. 2.
58. The values are widely spread around the mean.

- 59. The second class is far more variable; its scores lie much further from the mean.
- 60. The mean rises by 5; the standard deviation is unchanged.
- 61. Both double.
- 62. Squaring removes the sign, so positive and negative deviations do not cancel to zero.
- 63. Range 8 and mean 14.
- 64. It uses every value, so a single extreme point does not define it.

SECTION G - SAMPLING AND STUDY DESIGN

- 65. The population is every member of the group of interest; the sample is the subset actually measured.
- 66. Split the population into groups and sample from each in proportion to its size.
- 67. Choose a random start, then take every kth member of an ordered list.
- 68. 30 boys.
- 69. People at work are unlikely to be included, so the sample is not representative.
- 70. It can be too costly, too slow, or destructive to measure every member.
- 71. Every member has an equal chance of selection, which reduces systematic bias.
- 72. An experiment assigns the treatment; an observational study records what already happens, so confounding is harder to rule out.
- 73. Otherwise any difference could be caused by the other changes instead.
- 74. Non-response bias: the 88% who did not reply may differ systematically from those who did.

SECTION H - CORRELATION AND LINES OF BEST FIT

- 75. A strong positive linear relationship.
- 76. A very weak negative linear relationship.
- 77. From -1 to 1 inclusive.
- 78. 19.
- 79. It extrapolates far beyond the data, where the relationship may not hold.
- 80. A confounding variable, hot weather, drives both; neither causes the other.
- 81. The difference between an observed value and the value the line predicts.
- 82. Points scattered with no consistent upward or downward trend.
- 83. Curvature, clusters or outliers in the relationship.
- 84. Yes. r measures only linear association, so a strong curved relationship can still give r near 0.

SECTION I - CONDITIONAL PROBABILITY AND TREE DIAGRAMS

- 85. $1/3$, from $(6/10) \times (5/9)$.
- 86. $8/15$.
- 87. $2/3$.
- 88. It also depends on how common the condition is; for a rare condition most positives can be false positives.
- 89. 0.8.
- 90. 0.1.
- 91. $P(A \text{ and } B)$ divided by $P(B)$, for $P(B)$ greater than 0.
- 92. 0.4.
- 93. $1/4$.
- 94. It shows how the second set of probabilities changes after each first outcome.

SECTION J - MIXED REVIEW AND READING REPORTS

- 95. 13.5.
- 96. Mode 9 and range 8.
- 97. 12.5.
- 98. 12.
- 99. Positively skewed; the upper half is more spread out than the lower half.
- 100. At 40, which is half the total.
- 101. The baseline risk; doubling a very small risk still leaves a very small risk.
- 102. A few very expensive properties pull it upwards; the median describes a typical price better.
- 103. Not confidently; the interval from 49% to 55% includes values below 50%.
- 104. Who collected the data and how, and what is being compared with what.

Review and next steps

Mark your work against the answer key, record your score for each section, then answer the three questions below. Bring this page to your next lesson so your tutor can start from what actually blocked you.

Which question did you find hardest, and what exactly stopped you?

Rewrite the corrected working for that question here.

Which section do you want to practise again before your next lesson?

Score record

SECTION	QUESTIONS	NUMBER CORRECT	PRACTISE AGAIN?
A. Summarising data	11		
B. Spread and displays	11		
C. Probability	11		
D. Combined events and two-way tables	11		
E. Interpreting and evaluating claims	10		
F. Standard deviation and variance	10		
G. Sampling and study design	10		
H. Correlation and lines of best fit	10		
I. Conditional probability and tree diagrams	10		
J. Mixed review and reading reports	10		
Total	104		

Keep going with one-to-one support.

This is an original Noble Educators topic sample, not a complete syllabus and not an official exam paper. Check the topic, level and any required methods with your school or tutor.

Book a free demo lesson at nobleeducators.com/contact