

Essential Maths Skills
for AS/A-level

Psychology Answers

Molly Marshall

1 Arithmetic and numerical computation

Expressions in decimal and standard form

Decimal form

Guided question (p.9)

1 Table A.1

Millions	Hundred thousands	Ten thousands	Thousands	Hundreds	Tens	Ones	and Decimal point	Tenths $\left(\frac{1}{10}\right)$	Hundredths $\left(\frac{1}{100}\right)$	Thousandths $\left(\frac{1}{1000}\right)$	Ten-thousandths $\left(\frac{1}{10000}\right)$	Hundred-thousandths $\left(\frac{1}{100000}\right)$	Millionths $\left(\frac{1}{1000000}\right)$
						0	.	4	3				
						0	.	6					
					6	0	.	0	5				
					3	0	.	6	6				
					2	2	.	0	7	2			

Practice questions (p.9)

- 2 C because $\frac{7}{100} = 0.07$
- 3 A because $\frac{550}{1000} = 0.550$ (or 0.55)
- 4 B because six hundred and sixty-six and six tenths = 666.6
- 5 B because seventy and sixty-nine thousandths = 70.0069

Standard form

Guided question (p.12)

1 Table A.2

Decimal form	Standard form	What calculation needs to be done
4200	4.2×10^3	Multiply by 1000
0.00085	8.5×10^{-4}	Divide by 10 000
15.5	1.55×10	Multiply by 10
0.00888	8.88×10^{-3}	Divide by 1000
5 500 000 000	5.5×10^9	Multiply by 1 000 000 000

Practice question (p.12)

- 2 a Group A silence scored 16.30 which is highest.
b Group C canteen scored 14.68 which is lowest.
c i Group A = 1.63×10
ii Group B = 1.605×10
iii Group C = 1.468×10

Ratios, fractions and percentages

Ratios

Guided question (p.14)

- 1 Henry planted 14 plants in total: 4 tomato plants and 10 strawberry plants.

Step 1: calculate the total plants as $4 + 10 = 14$

Step 2: the ratio of strawberry plants to plants is 10 : 14

Step 3: both are divisible by 2 so the simplified ratio is ratio 5 : 7

Tom received 120 votes and Harry received 200 votes.

Step 1: calculate the total votes as $120 + 200 = 320$

Step 2: the ratio of Tom's votes to the total votes is 120 : 320

Step 3: to simplify, both sides can be divided by 10 so the ratio is 12 : 32

Step 4: as a ratio of 20 : 32 both sides can be divided by 4 so ratio is 3 : 8

The final simplified ratio of Tom's votes to total votes is 3 : 8.

Practice question (p.14)

- | | |
|--|---|
| 2 a 350 women + 140 men = 490 participants | c Male selfie : female selfie
18 : 45
2 : 5 |
| b Female : male
350 : 140
35 : 14
5 : 2 | d Female talk : male talk
240 : 80
3 : 1 |

Fractions

Guided question (p.16)

- 1 b Step 1: we know that 50% of the men have a type B personality.

Step 2: 50% can be written as $\frac{50}{100}$

Step 3: simplify the fraction as $\frac{1}{2}$

Step 4: as a fraction, $\frac{1}{2}$ (half) the men had a type B personality.

Practice question (p.16)

- 2 a B 5 : 1 is correct because 50 : 10 simplifies to 5 : 1
- b A 5 : 4 is correct because 50 : 40 simplifies to 5 : 4
- c i $\frac{5}{12}$
- ii $\frac{4}{12}$ which simplifies to $\frac{1}{3}$
- iii $\frac{2}{12}$ which simplifies to $\frac{1}{6}$

Percentages (%)

Guided question (p.19)

- 1 Step 1: 85 said YES. The total sample was 240.
- Step 2: $85 \div 240 = 0.35416$
- Step 3: $0.35416 \times 100 = 35.416$
- Step 4: round up to two decimal places = 35.42%
- Answer: 35.42% of parents said their 7-year-old son had a mobile phone.

Practice question (p.19)

- 2 a 165 males were observed.
- b 396 females were observed.
- c $165 \text{ (male participants)} \div 561 \text{ (total participants)} = 0.29411$
 $0.29411 \times 100 = 29.411$ rounded down to 29.41%
- d $396 \text{ (female participants)} \div 561 \text{ (total participants)} = 0.70588$
 $0.70588 \times 100 = 70.588$ rounded up to 70.59%
- e $63 \text{ (took selfie)} \div 561 \text{ (total participants)} = 0.11229$
 $0.11229 \times 100 = 11.229$ rounded up to 11.23%
- f $130 \text{ (used phone to text)} \div 561 \text{ (total participants)} = 0.23172$
 $0.23172 \times 100 = 23.172$ rounded down to 23.17%
- g $322 \text{ (used phone to talk)} \div 561 \text{ (total participants)} = 0.57397$
 $0.57397 \times 100 = 57.397$ rounded up to 57.40%

Estimating results

Guided questions (p.20)

- 1** Step 1: round up 185 participants to 200 (the nearest hundred).
Step 2: round down average score to 4.
Step 3: multiply $200 \times 4 = 800$
Exact answer: $185 \times 4.3 = 795.5$
- 2** Step 1: round down 130 participants to 100 (the nearest hundred).
Step 2: round up time needed for each participant from 17 to 20 minutes (nearest ten).
Step 3: multiply $100 \times 20 = 2000$ minutes
Step 4: divide by the number of minutes in an hour
 $2000 \div 60 = 33.333$
Step 5: round up to nearest hour = 34 hours
Exact answer: $130 \times 17 = 2210$ minutes
 $2210 \div 60 = 36.833$ hours
 $= 37$ hours
- 3** Step 1: round up the number of students to 60 000 (nearest thousand)
Step 2: round up the percentage who will be awarded an A grade to 20% (nearest whole number)
Step 3: calculate 20% of 60 000 ($60\,000 \div 100 \times 20$) as 12 000 students
Exact answer: $18.9 \div 100 = 0.189$
 $0.189 \times 58\,000 = 10\,962$ students
-

Practice question (p.21)

- 4 a** 'Self rating' average calculated as
191 rounded up to nearest hundred = 200; 18 participants rounded up to nearest 10s = 20, so estimated average = $200 \div 20 = 10$
Exact answer: $= \frac{191}{18} = 10.61111\dots$ so 10.61 to two decimal places
- b** 'Friend rating' average calculated as
282 rounded up to nearest hundred = 300; 18 participants rounded up to nearest 10s = 20, so estimated average = $300 \div 20 = 15$
Exact answer: $= \frac{282}{18} = 15.6666\dots$ so 15.67 to two decimal places

2 Handling data

Significant figures

Rounding significant figures

Guided questions (p.23)

- 1** Step 1: identify the third s.f. as 0.83**5**6.
Step 2: round UP to 0.836.
 - 2** Step 1: identify the second s.f. as 1.4**6**4.
Step 2: round UP to 1.5.
 - 3** Step 1: identify the third s.f. as 88.3**6**3.
Step 2: the following number 6 is greater than 5 so round UP to 88.4.
 - 4** Step 1: identify the third s.f. as 0.05**6**87.
Step 2: the following number 8 is greater than 5, so round up to 0.057.
 - 5** Step 1: identify the fourth s.f. as 33.3**3**3.
Step 2: the following number 3 is less than 5 so the answer is 33.33.
-

Practice questions (p.24)

- 6** 0.700 as two significant figures is 0.70
- 7** 5.548 as four significant figures is 5.548
- 8** 6.683 as three significant figures is 6.68
- 9** 0.564 as two significant figures is 0.56

Levels of measurement

Nominal level data

Guided question (p.24)

- 1 a** apples to pears
Step 1: 50 people chose an apple and 15 people chose a pear so the ratio is 50 : 15
Step 2: 50 and 15 are both divisible by 5, so the simplified ratio is 10 : 3
- b** apples to bananas
Step 1: 50 people chose an apple and 35 people chose a banana so the ratio is 50 : 35
Step 2: 50 and 35 are both divisible by 5 so the ratio is 10 : 7

- c** Step 1: 15 people chose a pear divided by the total number of observations of 100 (50 apples + 35 bananas + 15 pears) so $\frac{15}{100} = 0.15$
 Step 2: multiply 0.15 by 100 = 15
 Step 3: round up the answer so the answer is 15% of people chose a pear
- d** Step 1: 35 people chose a banana divided by total observations of 100 so $\frac{35}{100} = 0.35$
 Step 2: multiply 0.35 by 100 = 35
 Step 3: round up the answer so 35% of people chose a banana

Practice question (p.25)

- 2 a** red roses to pink roses

Step 1: 18 people chose red roses and 22 people chose pink roses so the ratio is 18:22
 Step 2: 18 and 22 are both divisible by 2, so the simplified ratio is 9:11

- b** yellow roses to pink roses

Step 1: 25 people chose yellow roses and 22 people chose pink roses so the ratio is 25:22

Step 2: there is no common denominator for 25 and 22 so the ratio is 25:22

- c** Step 1: 22 people chose pink roses divided by the total number of observations of 65 (18 red roses + 25 yellow roses + 22 pink roses) = $22 \div 65 = 0.33846$

Step 2: multiply 0.33846 by 100 = 33.846

Step 3: round up the answer to two decimal places so .846 is rounded up to .85 and the answer is 33.85%

- d** Step 1: 18 people chose red roses divided by the total number of observations of 65 (18 red roses + 25 yellow roses + 22 pink roses) = $18 \div 65 = 0.27692$

Step 2: multiply 0.27692 by 100 = 27.692

Step 3: round down the answer to two decimal places so .692 is rounded to .69 and the answer is 27.69%

Ordinal level data

Practice questions (p.25)

- 1 a** Scores: lowest to highest

Table A.3

1	2	3	4	5	6	7	8
22	28	33	34	37	41	45	48

- b** $22 + 45 + 34 + 33 + 48 + 28 + 37 + 41 = 288$

- c** Step 1: 22 is the lowest mark. The examination was marked out of 50.

Step 2: 22 divided by 50 is 0.44.

Step 3: multiply 0.44 by 100 so 44 becomes 44% (to 2 d.p.)

- d** Step 1: 48 is the lowest mark. The examination was marked out of 50.
Step 2: 48 divided by 50 is 0.96
Step 3: multiply 0.96 by 100 so 96 becomes 96% (to 2 d.p.)

Interval level data

Practice question (p.26)

- 1 a** Step 1: $69 \text{ seconds} \times 2 = 138 \text{ seconds}$
Step 2: 138 seconds divided by 60 = 2 minutes and (remainder) 18 seconds
- b** Step 1: $115 \text{ seconds} \times 2 = 230 \text{ seconds}$
Step 2: 230 seconds divided by 60 = 3 minutes and (remainder) 50 seconds
- c** Step 1: Maddie time (115 s) minus Sam time (69 s) = 46 secs
Step 2: $\frac{46\text{s}}{69\text{s}} = 0.66666\dots$
Step 3: multiply 0.66666 by 100 = 66.666
Step 4: round up to two decimal places. Thus Maddie took 66.66% longer than Sam to run 400 metres.

Examples of levels of measurement

Guided questions (p.27)

- 4** This is ordinal level data because the scores for each group on the psychology test can be put in order (ranked) lowest to highest.
- 5** The sponsors can be counted so this is nominal level data (frequency level) that can only be counted. So the measurement is frequency of occurrence.
- 6** This is interval level data because time to run 100 metres is measured on a fixed scale: hours, minutes, seconds, etc.

Practice question (p.27)

- 7 a** The time taken for participants to react to a sound stimulus is interval level data. It is measured on a fixed scale.
- b** An observation in a supermarket looking at how many males and how many females buy a lottery ticket is nominal level data. It is counted as a frequency.
- c** The number of participants that were obedient and administered a 450 volt electric shock is nominal level data. It is counted as a frequency.
- d** One point awarded for each correct answer is ordinal level data. The scores can be ranked lowest to highest.
- e** Counting the number of problems the students can solve correctly in 5 minutes is ordinal level data. The scores can be ranked.
- f** Having participants self-report the length of time (in hours and minutes) they slept the previous night is interval level data because time is measured on a fixed scale.

Measures of central tendency: mean, median and mode

Guided question (p.30)

- 1 d Use the median because the data is ordinal (the scores from the students on the spelling test can be ranked) and the sample is small.
- e Use the median because the data is ordinal (the number of psychology questions solved by each group can be ranked) and the sample is small.
- f Use the mode because the data is nominal and the film types fit into only one of the six categories.

Practice questions (p.31)

- 2 a Scores: lowest to highest

Table A.4

Without pictures	9	10	10	12	13	14	15	16	17
With pictures	10	11	13	14	15	16	16	16	17

- b i Median score without pictures is 13
ii Median score with pictures is 15
- c i Modal score without pictures is 10
ii Modal score with pictures is 16
- 3 a Use the mode because the data is nominal and the observed behaviour (choice of cake) fits into only one of the five categories.
- b Use the median because the data is ordinal (the two sets of scores can be ranked) and the sample of 10 students is small.
- c Use the median OR the arithmetic mean. Median because the data is ordinal (the scores from each group can be ranked). Mean because the data is ordinal, the mean takes account of the value of all the scores, and the sample is large.
- d Use the mean because the data is interval level (time taken to run 100 m) and the mean takes account of the value of all the scores.
- e Use the mode because the data is nominal and the observed car colour fits into only one of the seven colour categories.

Calculating the arithmetic mean

Guided questions (p.33)

- 1 $112 \div 5 = 22.4$
- 2 $892 \div 3 = 297.333\dots$ rounded to two decimal places is 297.33
- 3 There are 5 scores so the mean $= \frac{140 + 45 + 90 + 63 + 22}{5} = \frac{360}{5} = 72$
- 4 There are 4 scores so the mean $= \frac{21.5 + 55.3 + 23.8 + 18.5}{4} = \frac{119.1}{4} = 29.775$ and rounded to two decimal places is 29.78

Practice question (p.34)

- 5** The total relaxation score before stroking the cat was 95. There are 8 scores, so the mean is $\frac{95}{8} = 11.875$ rounded UP to two decimal places = 11.88

The total relaxation score after stroking the cat was 116. There are 8 scores, so the mean is $\frac{116}{8} = 14.50$ to two decimal places

Principles of sampling

Practice questions (p.36)

- 1 a** This is a random sample as the name of every student who is registered with the college was included and every one of them was equally likely to be selected.
- b** This is a systematic sample. The ratio of male and female participants in the sample is the same as the ratio of males and females in the target population.
- 2** $\frac{30}{250} \times 100 = 12\%$
- 3** The percentage which withdrew = $\frac{4}{50} \times 100$ which is 8%.
So the percentage of participants which remained = $100 - 8 = 92\%$.
- 4** $\frac{4}{100} \times 35 = 14$ participants all attend the same yoga class
- 5 a** School A: the ratio is 50m : 75f simplified as 10m : 15f; simplified as 2m : 3f
- b** School B: the ratio is 100m : 200f simplified as 10m : 20f; simplified as 1m : 2f
- c** Use a stratified sample to select a representative sample from School A.
- d** There are more female students than male students in School A: 50 males and 75 females. A stratified sample is needed to collect an unbiased sample so that the ratio of males to females in the sample is the same as in the population. In School A, the ratio of male to female students is 2 : 3, so for every 2 males in the sample there must be 3 females.
- e** To select a random sample from School B you would put all 100 male names and all 200 female names on individual slips of paper, fold them up and put them in a box and then pick out 25 slips from the box.
- 6 a** $25 + 40 + 15 = 80$ participants
- b** younger participants aged 18–20 to older participants aged 41–60
ratio = 25 : 15; simplified as 5 : 3
- c** younger participants aged 18–20 to participants over the age of 21
ratio = 25 : 55; simplified as 5 : 11
- d i** aged 18–20
 $\frac{25}{80} \times 100 = 31.25\%$
- ii** aged 21–40
 $\frac{40}{80} \times 100 = 50\%$
- iii** aged 41–60
 $\frac{15}{80} \times 100 = 18.75\%$

Simple probability

Guided question (p.38)

- 1 Throwing a 3 is an event with one favourable outcome (a throw of 3) and the total number of possible outcomes is six (1, 2, 3, 4, 5, or 6). Thus, the probability of throwing a 3 is $\frac{1}{6}$.

Practice questions (p.38)

- 2 Pulling out the black marble is the only favourable outcome. The total number of possible outcomes is eight (because there are eight marbles). Thus, the probability of pulling out the black one is $\frac{1}{8}$.
- 3 There are two favourable outcomes (your 2 tickets) and the total number of possible outcomes is 20 (because there are 20 tickets). Thus, the probability of you winning a prize is $\frac{2}{20}$ or simplified as $\frac{1}{10}$.
- 4 Pulling out a blue fish is the only favourable outcome. The total number of possible outcomes is 35 (because there are 35 fish). Thus, the probability of pulling out a blue fish is $\frac{20}{35}$. This can be calculated as a percentage as 57.14%.

Statistical significance

Guided questions (p.41)

- 1 Step 1: there are up to 10 chances in 100 that the result is due to chance.
Step 2: $100 - 10 = 90\%$
- 2 Step 1: there are up to 5 chances in 100 that the result is due to chance.
Step 2: $100 - 5 = 95\%$
- 3 Step 1: there is 1 chance in 100 that the result is due to chance.
Step 2: $100 - 1 = 99\%$

Practice question (p.42)

- 4 a If the results are significant at $p \leq 0.05$, then the researcher can be at least 95% confident that the difference in the time taken to sort the cards between the group with and without an audience is not a chance effect. The researcher can confidently state that there is only a less than or equal to 5% chance that having an audience did not cause the participants to sort the cards faster.
- b If the results are significant at $p \leq 0.01$, then the researcher can be at least 99% confident that the difference in the time taken to sort the cards between the group with and without an audience is not a chance effect. The researcher can confidently state that there is only a less than or equal to 1% chance that having an audience did not cause the participants to sort the cards faster.
- c If the results are significant at $p \leq 0.05$, then the researcher can reject the null hypothesis because the analysis shows that there is only a less than or equal to 5% probability that the difference in the time taken to sort the cards between the group with and without an audience is a chance effect. The researcher can be at least 95% confident that having an audience did cause the participants to sort the cards faster.

Measures of dispersion

Practice questions (p.44)

- 2 a If the highest score is 120 and the lowest score is 40, **the range is $120 - 40 = 80$**
b If the highest score is 556 and the lowest score is 331, **the range is $556 - 331 = 225$**
c If the highest score is 95.5 and the lowest score is 35.25, **the range is $95.5 - 35.25 = 60.25$**
d If the lowest score is 1.25 and the highest score is 5.5, **the range is $5.5 - 1.25 = 4.25$**
e If the lowest score is 1 minute 30 seconds and the highest score is 3 minutes 10 seconds
Step 1: convert highest score of 3 minutes 10 seconds to seconds = 190 seconds
Step 2: convert lowest score of 1 minute 30 seconds to seconds = 90 seconds.
Step 3: subtract lowest score from highest score 190 seconds – 90 seconds
= 100 seconds
Step 4: convert back to minutes and seconds, **so range is 1 minute 40 seconds**
- 3 In the first mathematics test the highest score is 46 and the lowest score is 25 **so range is $46 - 25 = 21$** .
In the second mathematics test the highest score is 42 and the lowest score is 27 **so range is $42 - 27 = 15$** .

Normal and skewed distributions

Normal distribution

Guided questions (p.45)

- 2 Step 1: what percentage of scores lie within 2SD of the mean in a normal distribution?
95% of the scores will be within $\pm 2SD$ from the mean.
Step 2: what percentage of scores lie within 2SD below the mean in a normal distribution?
Half of the 95% will be below the mean (the other half will be above the mean). $95\% \div 2 = 47.5\%$, thus 47.5% of 200 which is 95 of the scores will lie within 2SD below the mean.

Practice questions (p.46)

- 3 68% of the scores will be $\pm 1SD$ from the mean. Half of the 68% will be below the mean (the other half will be above the mean). So because there are 500 scores 34% of the scores will lie within 1SD below the mean.
34% of 500, is $\left(\frac{500}{100}\right) \times 34$ so 170 scores will lie within 1SD below the mean.
- 4 $\frac{100}{34} \times 100 = 294.117$ so, rounded up to a whole number, there were 294 scores in the data set
- 5 No, because in a normally distributed data set only 34% of scores will lie between the mean and 1SD above the mean. 34% of 200 scores is 68 scores, but in this data set 120 scores lie between the mean and 1SD above the mean so the data is not normally distributed.

Skewed distributions

Practice question (p.48)

- 2 a The data is skewed because 15 of the 25 scores fall below the mean score.
b The high scores (outliers) of 85 have pulled the mean score 'upwards'.
c The mean (of 57.4) is to the left of the highest point.
d 85 is the modal score because 5 students scored 85
e The range of exam scores is 85 (the highest) minus 30 (the lowest) which is 55.

Standard deviation

Calculating the standard deviation

Practice question (p.52)

2 Table A.5

Participant	Number of words remembered	Score minus mean ($x - \bar{x}$)	Score minus mean squared ($(x - \bar{x})^2$)
1	21	$21 - 16 = +5$	25
2	22	$22 - 16 = +6$	36
3	18	$18 - 16 = +2$	4
4	12	$12 - 16 = -4$	16
5	11	$11 - 16 = -5$	25
6	18	$18 - 16 = 2$	4
7	15	$15 - 16 = -1$	1
8	16	$16 - 16 = 0$	0
9	12	$12 - 16 = -4$	16
10	15	$15 - 16 = -1$	1
Total score	160		
Mean score	16		
Sum of squared differences			128

How is the standard deviation useful?

Practice questions (p.54)

- 2 We can conclude that younger people have better memories than older people because the *difference in mean memory scores* shows that the younger people scored, on average, 1.8 ($15.4 - 13.6$) higher than older people. Also, the *memory scores of the younger people are less varied* (dispersed/spread out) compared to the scores of the older people. *The lower standard deviation* in the young group suggests that individual memory scores are similar to each other as, *with a small standard*

deviation of 2.6, all are quite close to the average score of 15.4. In contrast, the standard deviation in the older group memory scores, at 5.5, is nearly double that of the variation in the younger group scores.

This answer would be awarded high marks. The conclusions in respect of both means and standard deviations are clearly presented. The answer demonstrates understanding of both the mean and the standard deviation and the use of data demonstrates understanding. The content in italics shows that the student can interpret data and understands the use of standard deviation.

3 Step 1: the variance is $\frac{2028}{n-1} = \frac{2028}{9} = 225.333\ldots$

Step 2: standard deviation is the square root of the variance

$\sqrt{225.333} = 15.0111$ or 15.01 to two decimal places

Selecting an appropriate statistical test

Practice questions (p.58)

- 5 The Mann–Whitney U test is a non–parametric test and is a test of the significance of the difference between two conditions suitable for use when an independent design has been used AND the level of data collected is ‘at least’ ordinal.
- 6 The Related T test is a parametric test and is a test of the significance of the difference between two conditions when a repeated measures design has been used AND the level of data collected is ‘at least’ ordinal AND the data is normally distributed.
- 7 The Spearman’s Rho (Rank Order) correlation coefficient is used when a correlation between two independent variables is being analysed. Spearman’s is a non-parametric test that calculates the correlation coefficient between ranked scores when both sets of scores (the age and the verbal ability scores) are ‘at least’ ordinal data (can be placed in rank order high to low or low to high).

8 a Table A.6

	High self-efficacy	Medium self-efficacy	Low self-efficacy
Good examination outcome			
Middling examination outcome			
Weak examination outcome			

- b The Chi-square test is appropriate because it is a test of significance of association which is used when nominal level data (frequency data) has been collected.

Using statistical tables to determine significance

Guided questions (p.60)

1 Table A.8

Number of pairs of scores	Level of significance	Critical value
10	$p \leq 0.05$	0.58
8	$p \leq 0.01$	0.76
5	$p \leq 0.05$	0.75
10	$p \leq 0.01$	0.71
7	$p \leq 0.05$	0.67
6	$p \leq 0.01$	0.83

2 Table A.9

Number of pairs of scores	Level of significance	Calculated correlation coefficient (R)	Critical value	Reject or retain null hypothesis
10	$p \leq 0.05$	0.55	0.58	Retain null because 0.55 is less than 0.58
8	$p \leq 0.01$	0.85	0.76	Reject null because 0.85 is greater than 0.76
7	$p \leq 0.05$	-0.84	0.67	Reject null because -0.84 is greater than 0.67*
9	$p \leq 0.05$	0.55	0.60	Retain null because 0.55 is less than 0.60
5	$p \leq 0.01$	-0.77	0.87	Retain null because -0.77 is less than 0.87

* -0.84 is greater than the table value of 0.67 because, when comparing the calculated correlation coefficient to the critical value in the table, you ignore the positive or negative sign.

Chi-square and degrees of freedom

Guided questions (p.63)

- 1
 - a 2×2 Chi-square $df = (2 - 1) \times (2 - 1) = 1 \times 1 = 1$
 - b 2×3 Chi-square $df = (2 - 1) \times (3 - 1) = 1 \times 2 = 2$
 - c 3×2 Chi-square $df = (3 - 1) \times (2 - 1) = 2 \times 1 = 2$
 - d 4×2 Chi-square $df = (4 - 1) \times (2 - 1) = 3 \times 1 = 3$
 - e 3×4 Chi-square $df = (3 - 1) \times (4 - 1) = 2 \times 3 = 6$
 - f 5×3 Chi-square $df = (5 - 1) \times (3 - 1) = 4 \times 2 = 8$

2 Table A.10

Significance level	Degrees of freedom	Calculated Chi-square	Significant
$p \leq 0.05$	3	10.5 greater than 7.82	No
$p \leq 0.05$	4	8.8 smaller than 9.49	Yes
$p \leq 0.01$	2	12.25 greater than 9.21	No
$p \leq 0.01$	3	9.5 smaller than 11.35	Yes

Practice question (p.64)

3 Table A.11

	Calculated Chi-square value	Degrees of freedom	Level of significance
a	15	1	$p \leq 0.05$
b	2.5	2	$p \leq 0.05$
c	10	3	$p \leq 0.01$
d	12	4	$p \leq 0.05$
e	5	1	$p \leq 0.01$

- a** The critical value at $p \leq 0.05$ for 1 df is 3.84, the calculated value is 15, which is greater than 3.84 so the result is not significant.
- b** The critical value at $p \leq 0.05$ for 2 df is 5.99, the calculated value is 2.5, which is less than 5.99 so the result is significant.
- c** The critical value at $p \leq 0.01$ for 3 df is 11.35, the calculated value is 10, which is less than 11.35 so the result is significant.
- d** The critical value at $p \leq 0.05$ for 4 df is 9.49, the calculated value is 12, which is greater than 9.49 so the result is not significant.
- e** The critical value at $p \leq 0.01$ for 2 df is 9.21, the calculated value is 5, which is less than 9.21 so the result is significant.

Differences between quantitative and qualitative data

Practice questions (p.66)

1 Possible answers:

Putting bags on seats

Putting coats on seats

Sitting in aisle seat to prevent access to window seat

2 Possible answers:

Waving arms

Tossing hair

Making rude gestures

Folding arms

Turning away

Touching others on arms

Hugging others

Difference between primary and secondary data

Practice questions (p.67)

- 1** This is primary data because the observer collected the data directly from the participants.
- 2** This is primary data because the researcher collected the data directly from the participants.
- 3** This is primary data because the researcher collected the data directly from the participants who answered the questionnaire.
- 4** This is secondary data because the researcher gathered the data from secondary sources.

Order of magnitude

Practice question (p.69)

- 2 a** Too much, because 1000 milligrams make 1 gram, and each box contains 10 000 milligrams which is 10 grams.
- b** Each box weighed one order of magnitude greater than required, because each box should have weighed 5000 milligrams (equivalent to 5 grams) but in fact weighed 10 000 milligrams (equivalent to 10 grams).

3 Algebra

Algebraic symbols

Practice question (p.71)

1	True or False
a $55 > 20$	T
b $5000 < 7000$	T
c $666 \neq 666$	F
d $500 \geq 250$	T
e $333 \leq 400$	T
f If $x = 8$ and $y = 9$, then $x + y = 20$	F
g $11 + 99 \neq 100$	T
h $33 + 22 > 10$	T
i $599 - 199 \sim 400$	T
j $55 \leq 65$	T
k $100 \gg 0.00001$	T

Substituting numerical values into algebraic equations

Practice question (p.72)

1 Table A.12

	Symbol	What the symbol means
a	x	Each value in the sample of scores
b	n	The number of participants
c	$\bar{x}$	The average score (the mean)
d	Σ	The sum of the scores (the total score)
e	$\geq$	Is equal to or greater than
f	$\leq$	Is equal to or less than

Substitution in the Chi-square formula

Practice question (p.75)

2 Table A.13

Category	<i>O</i> (observed frequency)	<i>E</i> (expected frequency)	<i>O</i> − <i>E</i>	(<i>O</i> − <i>E</i>) ²	(<i>O</i> − <i>E</i>) ² ÷ <i>E</i>
Always eat five fruit/ veg and had ‘flu	60	50	10	100	$\frac{100}{50} = 2$
Never eat five fruit/veg and had ‘flu	70	50	20	400	$\frac{400}{50} = 8$
Always eat five fruit/ veg and did not have ‘flu	40	50	−10	100	$\frac{100}{50} = 2$
Never eat five fruit/veg and did not have ‘flu	30	50	−20	400	$\frac{400}{50} = 8$
Total	200	200			2 + 8 + 2 + 8 = 20

4 Graphs

Frequency tables, diagrams, bar charts and histograms

Practice question (p.81)

2 a Draw a bar chart:

- Both axes must be labelled correctly.
 - Your chart must have an informative title.
 - The y-axis must have an appropriate scale.
 - There must be gaps between the bars.
 - The bars must be plotted reasonably accurately.
- b The data is displayed on a bar chart because car colours are discrete (nominal) data and a bar chart is the appropriate type of graph on which to show this type of data.
- c The data is nominal level data.
- d The mode is the appropriate measure of central tendency and the mode will show the most frequent car colour.

Scatter diagrams

Plot two variables from an investigation

Practice question (p.83)

1 a The mathematics examination marks are ordinal level data.

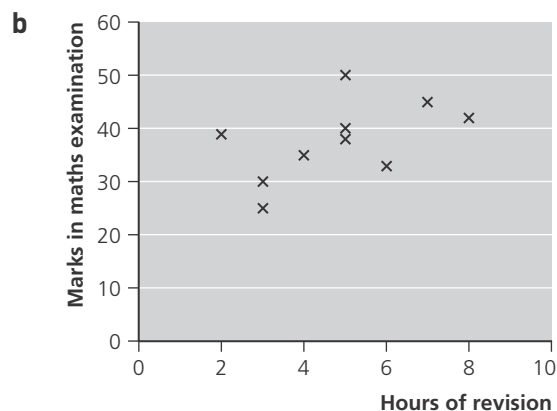


Figure A.1 Relationship between revising and marks in maths examination

- c The scatter diagram shows no correlation between exam marks and number of hours of revision.

Exam-style questions

1 a Table A.14

Participant	Attitude score	
	Male	Female
1	5	6
2	6	6
3	4	5
4	4	5
5	5	6
6	5	4
7	7	7
8	7	8
9	4	9
10	3	7
Mean rating	5.0	6.3
Mode rating	4 and 5	6

The mode of the male attitude scores is 4 and 5; the mode of the female attitude score is 6.

2 marks awarded for 2 correct answers

1 mark awarded if only 1 answer is correct

- b** The mode is not very useful because it does not take into account the value of the scores and tells us nothing about any other scores in the data set.

1 mark awarded for an accurate answer

- c** Answer 1: One measure of dispersion that could be used is the range, the difference between the lowest and highest attitude score.

Answer 2: One measure of dispersion that could be used is the standard deviation to show how the scores are distributed around the mean attitude score.

1 mark awarded for either of these accurate answers

- d** Answer 1: A Mann–Whitney U test could be used.

Answer 2: An Unrelated T test could be used.

1 mark awarded for either of these accurate answers

- e** Answer 1: The Mann–Whitney U test is appropriate because an independent design is used (males in one condition, females in the other condition), the data collected is ordinal level data and the research is testing the difference in male and female attitudes to the planned housing development.

Answer 2: The Unrelated T test is appropriate because an independent design is used (males in one condition, females in the other condition), the data collected is ordinal level data, there are no extremely high or low scores so the data looks to be normally distributed, and the research is testing the difference in male and female attitudes to the planned housing development.

3 marks awarded:

1 mark for justification based on design (the independent design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

- f** A bar chart should be drawn, with both axes labelled and an appropriate title.

3 marks awarded:

Marks are lost if there are no gaps between the bars, and/or axes are not labelled, and/or there is no title

- 2 a** Answer 1: A Wilcoxon matched pairs signed ranks test could be used.

Answer 2: A Related T test could be used.

1 mark awarded for either of these accurate answers

- b** The Wilcoxon matched pairs signed ranks test is appropriate because a repeated measures design is used (same participants in each condition), the data collected is ordinal level data and the research is testing the difference in memory for familiar objects compared to unfamiliar objects.

3 marks awarded:

1 mark for justification based on design (the repeated measures design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

- c** A bar chart should be drawn, with both axes labelled and an appropriate title.

3 marks awarded:

Marks are lost if there are no gaps between the bars, and/or axes are not labelled, and/or there is no title

- 3 a** Step 1: calculate the total time taken as $120 + 65 + 45 + 33 + 28 + 22 + 15 + 9 + 5 = 342$.

Step 2: divide the total by the number of participants as $\frac{342}{9} = 38$

The mean time taken for the chicken to ring the bell is 38 seconds.

2 marks awarded for accurate mean and calculations shown

1 mark awarded for accurate mean but no calculations shown

- b** C is the correct answer. The 1st time is 120 seconds, the 9th time is 5 seconds, thus a ratio of 4:1.

1 mark awarded for the correct answer

- c** Range is calculated as the slowest time minus the fastest time

$120 - 5$

$= 115$ seconds

1 mark awarded for the correct answer

- d** The level of learning is interval level data, as time taken for the chicken to ring the bell is measured on a fixed scale: hours, minutes and seconds.

1 mark awarded for the correct answer

- 4 a** One conclusion is that people are almost twice as likely to laugh out loud at jokes if they are with other people, than if they are alone. The mean 'laugh aloud' in the group condition is almost twice that in the alone condition (9.5 compared to 5.5).

Also, the standard deviation in the group condition is lower for both 'laugh aloud' and 'smiled' (1.25; 2.5) than in the alone condition (3.6 and 4.85) which suggests that there are individual differences in whether people find jokes funny when they listen to the joke on their own. However, it could also be concluded that people laugh at jokes just because other people do, rather than because they find the joke funny.

6 marks awarded:

This is a strong answer. The answer draws a reasonable conclusion based on the mean, and goes on to draw two appropriate conclusions based on the standard deviations.

- b** The level of data is ordinal because the researcher counted the number of times each of the participants laughed aloud or smiled.

1 mark awarded for the correct answer

- c** Answer 1: A Mann–Whitney U test could be used because an independent design is used (watched the film in a group or watched the film alone), the data collected is ordinal level data and the research is testing the difference between laughing aloud and smiling in the group or alone condition.

4 marks awarded:

1 mark for identifying an appropriate test

1 mark for justification based on design (the independent design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

Answer 2: An Unrelated T test could be used because an independent design is used (watched the film in a group or watched the film alone), the data collected is ordinal level data and the research is testing the difference between laughing aloud and smiling in the group or alone condition.

4 marks awarded:

1 mark for identifying an appropriate test

1 mark for justification based on design (the independent design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

- d** Standard deviation is a useful measure of dispersion because it shows how closely the scores cluster round the mean. A low standard deviation shows that most of the scores are 'near' the average score and a high standard deviation suggests that there may be anomalous scores (outliers) that are much higher or lower than the mean.

3 marks awarded:

1 mark for demonstrating understanding of standard deviation

2 marks for giving a clear and accurate explanation of how the standard deviation is useful

- e** If the results are significant at $p \leq 0.05$ it means that the researcher can be at least 95% confident that being in a group did influence whether people laughed or smiled as there is only, at most, a 5% probability that the difference between the number of times they laughed/smiled is caused by random chance.

2 marks awarded:

1 mark for giving a correct interpretation of $p \leq 0.05$, and 1 mark for giving an accurate expansion to demonstrate a thorough understanding

- 5 a** Answer 1: A measure of central tendency could be the arithmetic mean, because the mean takes account of the value of each of the scores and there are no extremely low or high memory scores to distort the mean.

Answer 2: A measure of central tendency could be the median, because there is only a small sample (10 scores) in each condition, the data can be ranked lowest to highest and the median will show the mid-point of the scores.

2 marks awarded for each appropriate answer:

1 mark for identifying an appropriate measure of central tendency and 1 mark for a clear and accurate justification

- b** A Mann–Whitney U test could be used because an independent design is used (the IV is gender of the participant), the data collected is ordinal level data (the number of male or female words recalled) and the research is testing the difference between male and female memory.

3 marks awarded:

1 mark for identifying an appropriate test

1 mark for justification based on design (the independent design)

1 mark for justification based on correct level of data

- c** If the results are not significant at $p \leq 0.05$ it means that the researcher can't be certain that people are more likely to remember gender related information because there is a more than 5% probability that the difference in the number of female (or male) words remembered is caused by random chance.

2 marks awarded:

1 mark for giving a correct interpretation of 'not significant at $p \leq 0.05$ ', and 1 mark for giving an accurate expansion demonstrating a thorough understanding

- 6 a** A scatter diagram should be drawn showing a positive relationship between the number of sessions attended and increasing confidence.

3 marks awarded:

1 mark for a scatter diagram

1 mark for labelling both axes

1 mark for an appropriate title

- b** The scatter diagram suggests that there is a positive correlation between confidence in mathematics ability and attending extra tuition sessions, because as the number of tuition sessions increases so does the average confidence level.

2 marks awarded:

1 mark for the accurate suggestion and 1 mark for the clear and in context explanation

- c** The level of data collected is ordinal: the number of tuition sessions attended and the self-report scores of mathematics confidence are numeric and can be ranked lowest to highest.

2 marks awarded if both variables are mentioned in the accurate answer

1 mark awarded if only one variable is mentioned

- d** A Spearman's Rho (Rank Order) correlation coefficient test could be used because this test will find the relationship between the number of tuition sessions attended and the final self-report mathematics confidence score, and both independent variables are ordinal level data.

3 marks awarded:

1 mark for identifying an appropriate test

1 mark for justification based on correlation

1 mark for justification based on correct level of data

- e** If the results are significant at $p \leq 0.01$ it means that the researcher can be at least 99% confident that there is a positive correlation between mathematics tuition and mathematics confidence so the null hypothesis should be rejected because there is only a less than or equal to 1% probability that the relationship between attending extra tuition and increasing mathematics confidence occurred by chance.

3 marks awarded:

1 mark for giving a correct interpretation of $p \leq 0.01$

2 marks for correctly stating that the null hypothesis should be rejected, and then clearly and accurately explaining why, demonstrating thorough understanding

- 7 a** The mean reaction time in the silent condition is 6 seconds.

1 mark awarded for the correct answer

- b** The participants who performed the task in silence had the fastest reaction time. Their average was 6 seconds compared to 9 seconds in the noisy condition.

2 marks awarded for the correct answer:

1 mark for identifying the correct condition and 1 mark for giving an accurate explanation

- c** The time in seconds is interval level data – interval level data is measured in fixed interval scale.

1 mark awarded for the correct answer

- d** Step 1: the ratio is 6 seconds (silent) to 9 seconds (noisy)

Step 2: simplify (divide both sides by 3) so the ratio of average reaction time in silent to noisy condition is 2 : 3

2 marks awarded:

1 mark for calculating the correct ratio as 6 : 9

1 mark for showing the simplification and final correct ratio

1 mark awarded if the answer only states the final ratio as 2 : 3

- e The Mann–Whitney U test is appropriate because the experiment has an independent design, there are different participants in the silent and noisy conditions AND the level of data (the reaction time scores) is interval level (which is at least ordinal) AND the test is appropriate when a difference between conditions is being tested.

4 marks awarded:

2 marks for the context of the study – the correct justification based on independent design

2 marks for the context of the study – correct justification based on level of data as interval

1 mark awarded each time if justification is not contextualised

- 8 a The order of scores for uncrowded is 3, 6, 11, 12, 12, 14, 15, 17, 18, 20 so median score is $12 + 14$ divided by $2 = 13$

The order of scores for crowded is 15, 15, 18, 18, 20, 20, 21, 22, 23, 25 so median score is $20 + 20$ divided by $2 = 20$

2 marks awarded: 1 mark for each correct answer

- b The modal score for uncrowded is 12 as this score occurs twice.

The modal scores for crowded are 15, 18, 20 as each of these scores appear twice .

2 marks awarded: 1 mark for each correct answer

2 marks awarded if both modal scores are fully accurate

1 mark awarded if one modal score is fully accurate

0 marks awarded if an inappropriate answer is given

NB: *If only one of the modal scores in the crowded condition is identified a mark is not awarded.*

- c Credit could be given for:

The mean because it is very sensitive and extracts most information from the scores; the data in this research suits the mean and there are no extreme scores in the crowded or uncrowded condition so data is not likely to be skewed, which favours the mean.

The median because in this small sample the median is not affected by extreme scores (as seen in uncrowded condition).

Answers are most likely to suggest the mean or the median and in an exam credit is unlikely to be awarded for answers which suggest the mode as a measure of central tendency because the mode is only appropriate when the data is nominal (frequency) level.

3 marks awarded for an appropriate suggestion that is justified with reference to the data in the study

2 marks awarded for an appropriate suggestion that is justified

1 mark awarded for an appropriate suggestion

- d Most likely answer: The standard deviation gives a better representation of the way the data is dispersed around the mean (the average cortisol levels) than the range. Unlike the range, the standard deviation takes account of the value of

every score and its distance from the mean, and the standard deviation gives an indication as to whether there are extremely high or low cortisol levels.

3 marks awarded for a correct identification plus a fully accurate description of an appropriate advantage

2 marks awarded for a correct identification plus a basic description of an appropriate advantage or full description of an appropriate advantage only

1 mark awarded for a correct identification

- e** Answer 1: A Wilcoxon matched pairs signed ranks test could be used.

The Wilcoxon matched pairs signed ranks test is appropriate because a repeated measures design is used (same participants in the uncrowded and crowded conditions), the data collected is interval level and the research is testing the difference between cortisol levels in uncrowded and crowded conditions.

4 marks awarded:

1 mark for correct identification of an appropriate test

1 mark for justification based on design (the repeated measures design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

Answer 2: A Related T test could be used.

The Related T test is appropriate because a repeated measures design is used (same participants in the uncrowded and crowded conditions), the data collected is interval level, the research is testing the difference between cortisol levels in uncrowded and crowded conditions and the data does not appear to have a skewed distribution.

4 marks awarded:

1 mark for correct identification of an appropriate test

1 mark for justification based on design (the repeated measures design)

1 mark for justification based on correct level of data

1 mark for justification based on research aim (in context)

- f** The hypothesis is a one-tailed hypothesis because it predicts that the mean cortisol level in the uncrowded condition will be lower than the mean cortisol level in the crowded condition (the direction of the effect of crowded/not crowded). A two-tailed hypothesis would only have predicted a difference in the means, but not the direction e.g. 'Levels of cortisol will be different after 30 minutes in an uncrowded environment than after 30 minutes in a crowded environment.'

2 marks awarded:

1 mark for correct identification of one-tailed hypothesis

1 mark for accurate explanation

NB: Only 2 marks are available so the student could save time and not add the accurate information about two-tailed hypothesis.

- g** If the difference between the cortisol levels in the uncrowded and crowded conditions is significant at $p \leq 0.05$ it means that the researcher can be at least 95% confident that it was the crowding (or not) that caused changed cortisol levels, so the null hypothesis should be rejected because there is only a less than or equal to 5% probability that the changes in cortisol levels were caused by random (chance) factors.

4 marks awarded:

2 marks for giving a correct interpretation of $p \leq 0.05$ in the context of the study

2 marks for correctly stating that the null hypothesis should be rejected, and then clearly and accurately explaining why, in the context of the study, demonstrating thorough understanding

- 9 a** $35 + 40 + 15 + 50 + 25 = 165$ ice creams

2 marks awarded for the working and final answer

1 mark awarded if the correct answer is given without working

- b** Ice-lolly is the most popular because 50 is more than each of the others.

1 mark awarded for the correct answer

- c** The mode is the correct measure of central tendency because the level of data is nominal – the different types of ice cream are the categories.

3 marks awarded if the measure of central tendency is identified correctly, the explanation is accurate and the answer is in context

1 mark awarded if the level of data is identified correctly but not explained

- d** $35 \text{ plain cones} + 40 \text{ '99' cones} = 75 \text{ cones.}$

The ratio is 75 cones : 50 ice-lollies

simplified as 3 : 2 (both sides divided by 25) so ratio = 3 cones : 2 ice-lollies

2 marks awarded for the correct answer

1 mark awarded if the ratio is expressed as 75 : 50 (not simplified)

- e** $\left(\frac{50}{165}\right) \times 100 = 30.303030\dots$ rounded to two decimal places = 30.30% of the sales are ice-lollies

2 marks awarded for the correct answer

1 mark awarded if the answer is not given correct to two decimal places

- 10 a i** dogs

$$\left(\frac{32}{280}\right) \times 100 = 11.43\%$$

- ii** spiders

$$\left(\frac{25}{280}\right) \times 100 = 8.93\%$$

- iii** wasps

$$\left(\frac{15}{280}\right) \times 100 = 5.36\%$$

iv cats

$$\left(\frac{8}{280}\right) \times 100 = 2.86\%$$

4 marks awarded: 1 mark for each correct answer

- b** $45\% = 126$ participants $\left(\frac{280}{100} \times 45 = 126\right)$ thus $280 - 126 = 154$ participants did not report fearing an animal or insect.

2 marks awarded if the full answer is correct

- c** Step 1: 126 reported fear : 154 did not report fear

Step 2: divide each side by 2 to simplify = 63 : 77

Step 3: divide each side by 7 (lowest common denominator) to simplify

The ratio is 9 : 11

9 (DID report fear):11 (DID NOT report fear)

2 marks awarded if the full answer is correct