20 Answers

20.1 Data Collection

1. (a) No, since $p(4\text{ Heads}) = \frac{1}{2^4} = 0.0625\ (>0.05)$

   (b) Yes, since $p(5\text{ Heads}) = \frac{1}{2^5} = 0.03125\ (<0.05)$

2. Discrete, Continuous, Discrete

3. (a) No. of wheels, etc. (b) Its weight, etc.

4. (a) Undertake a survey (b) Secondary data (c) May not be reliable

5. (a) Sample too small; samples may not be independent of each other
   (b) Easier and cheaper

6. Too small a sample, and not representative (as they are in the same family).

7. Nobody under age 18 interviewed; too many males (particularly in the 26-40, 41-64 age groups)

8. You would need a large sample to have proof!
   Indeed, for proof, you would need to survey all women.

9. (a) 600 (b) (i) 72% (ii) 88%
   (c) The problem is not the sample size, but the large number of non-respondents.
   (d) Undertake interviews (either face to face or by telephone).

10. (a) (i) Label 001 to 500 all pupils, and take random digits, three at a time, ignoring those over 500 or repeats.
    (ii) Choose a number between 01 and 50 at random, say x; then choose $x,\ x+50,\ x+100,\ \ldots\ x+450$ as the sample.
    (b) (i) BOY/GIRL or by age groups
    (ii) You are more likely to get a representative sample.

11. (a) 220 (b) 180 (c) 180, since it comes from a larger sample
    (d) You would need to add a species of fish not present at the moment and then apply the same techniques.

12. (a) Find the number of woodlice in similar areas of the garden, but some areas being damp and others not.
    (b) Survey no. of woodlice in dark, damp conditions and in light damp conditions.

13. (a) Method 2 would include pupils from all year groups.
    (b) Method 1 would only sample pupils who arrive early.
    Method 3 might not have enough pupils from some year groups.
    (c) Question is posed in a negative way.
    (d) e.g. "Do you think that school uniform should be worn?"

14. (a) (i) Discrete (ii) Qualitative (iii) Continuous (iv) Qualitative
    (b) (i) (i) and (iii) (ii) (ii) and (iv)
    (c) Scatter plot (d) Height and weight

15. (a) Only people travelling by train would be sampled.
    Sample would consist mainly of commuters on their way home - not representative.
    'First 100' is not a representative sample.
    (b) Use electoral register as sampling frame, and choose sample randomly.
    (c) (i) It is not possible to test the lifetime of all batteries.
    (ii) Too expensive to take a census
    (iii) Check for quality control at regular intervals.

16. (a) e.g. "How long do you think the interval should last?"
    (b) Provide a stamped, addressed envelope; offer an incentive such as reduction in price of a ticket to those who reply.
    (c) All ticket holders numbered, and 20 chosen (without replacement) at random.
    (d) (i) Choose one of the numbers 1 - 5 at random, and then every 5th person after that.
    (ii) $\frac{1}{5}$
    (e) 8
    (f) Random sampling, as it will not necessarily be representative.

17. (a) Temperature (as this could explain the difference in times)
    (b) Typing times (as these are potentially dependent on temperature)
    (c) Roughly equal numbers and similar typing skills in each group.

18. (a) Pupils of different gender will have different spending patterns.
    (b) (i) Different spending patterns.
    (ii) Number all boys in sampling frame, and choose 20 at random.

20.2 Presentation of Data

1. (a) 40 (b) 30 (c) 180 (d) 9 (e) £7500

2. (a) about 260 (b) about 10 000 (c) they need less

3. (a) 1 800 000 (b) UK (c) 200 000
    (d) Far fewer teachers in Somalia than in Vietnam.

4. (a) 29 (b) 6 (c) 6

5. (a) 370 000 (million) (b) 15 days (c) very large population

6. (a) £8562

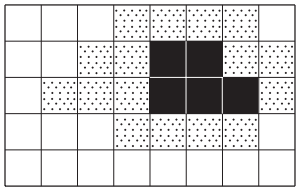
   (b) Each disc for Fairplan is worth £7132, which is far less than £8562.

7. (a) 6 (b) 36 (c) 120 (d) 15%, 25%, 10%, 30%, 5%, 15%

8. (a) 65 - 70 (b) (i) 10-15 (ii) about 14% (c) about 6.5%

   (d) e.g. In India, very few people live to over 70; in UK, significantly more females than males live beyond 70.

9. (a)



   (b) Towards the top right corner, as there are more pupils in that area.

10. (a) (i) Under 5% (ii) 5–10%

    (b) Risk is greater for men, as there is more of the darker shading in total on the diagram.

    (c) (i) Risk is higher with age

    (ii) Risk is less for non-smokers.

11. (a)

1	5 6 8
2	1 2 4 4 6 8 9
3	1 1 2 5 6 7 8
4	0 1 2 6 9
5	1 2 7

(b) (i) 42 (ii) 32

12. (a)

1	5 7 8 8 8 9 9
2	0 1 1 1 1 2 3 3 3 4 4 5 6 7 7 8
3	1 2 3 4 5 5 6 7
4	0 3 5 5 8
5	1 7 8
6	1 2 3 5 5 9
7	1 3
8	2

13.

			17	0 4 9 9
			18	2 3 5
			19	0 5 5
2	3	3	20	1 2
		5	21	0
		1	22	
		7	23	
0	6	9	24	0
	3	4	25	
2	6	6	26	3
		3	27	
	3	8	28	

For Volume 1, the data is more spread out than for Volume 2, and it has a higher average value.

14. (a) 181°, 93°, 51°, 35° (b) 3.14 cm
(d) Total decrease in attendance, but Division 1's share of attendance has increased.

15. (a) 98 (b) $\frac{37}{98}$ (c) $r^2 = 24^2 \times \frac{31}{39}$; $r \approx 21.4$ mm
(e) overall reduction with time, particularly at 'take off'
(f) multiple bar chart – gives better representation of each category, and overall trends.

16. Vertical axis does not start at zero; hence not in proportion to % values.

17. (a) The vertical axis starts at 5 which gives a misleading impression of the overall trends; also the 'Western Europe' group looks similar to the other groups whereas it is in fact much larger.
(b) 1985 (c) 7 (million) (d) 1989

18. (a) 1983 (b) Just over 6 bn Fr.
(c) All improving with SNCF continuing to be more productive than DB, and DB more productive than BR.
(d) The vertical scale does not start at zero, and hence exaggerates productivity.

20.3 Measures of Location and Spread : Box and Whisker Plots

1. (a) A discrete variable has a finite value (e.g. the number of matches in a box). A continuous variable is not finite and depends on the accuracy of measurement used (e.g. the heights of pupils in a class)

(b)

0	3	5	7	8	9	9	9	9						
1	0	0	0	0	1	2	3	3	4	5	5	6	6	7
2	0	1	2	3	4									

(c) Easy to construct: shows main features of distribution.

(d) Range 3 to 24; median = 12, LQ = 9, UQ = 16

2. 20th percentile $\Rightarrow$ 24th value, i.e. 1.8
80th percentile $\Rightarrow$ 96th value, i.e. 8.0
} range = 6.2
The true range (0 to 32) includes outliers, which are not really relevant.

3. (a) (i) 24.7 to 26.4 i.e. 1.7 (ii) $25.3 - 24.6 = 0.7$
(iii) It is far from the other data values.
(b) Firm C – as it is the only one to meet the requirements.

4. (a) 41 (b) 41 (c) 53 (d) Overall, he scored lower

5. (a)

3	6	7	7	7						
4	2	2	6	8	9					
5	1	1	3	3	4	5	6	7	7	8
6	0	2	3	4	8					
7	0									

(b) (i) 53 (ii) 37
(iii) $70 - 36 = 34$
(c) The median better represents the average.

6. (a)

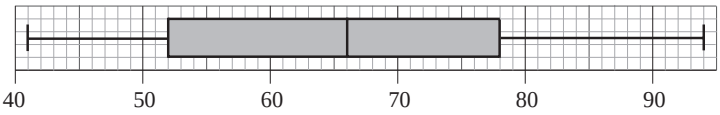
4	1	2	3	4	4	6	7	7	8	8
5	0	2	2	2	3	4	6	7	8	8
6	0	2	3	3	6	6	7	7	8	
7	0	0	2	2	4	4	6	7	8	8	8
8	0	1	2	5	5	6	6	7			
9	2	3	4								

(b) 66 ; 52, 78

(d) (i) Stem and leaf is easy to construct.

(ii) Box plot quickly shows you the main features.

(c)



Miles travelled by car each week

7. (a) 2, 3, 2, 5, 2, 1

(b) (i) $\frac{2}{15}$ (ii) $\frac{8}{15}$

8. (a) 25 30 30 31 33 34 36 36 37 37 38 38 38 40 41 43 43 48 55

LQ Median UQ

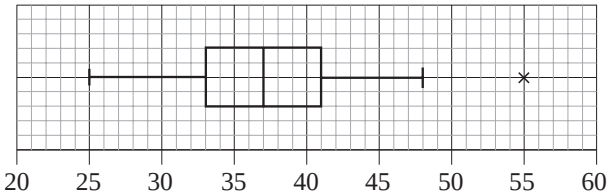
Hence IQR = 41 – 33 = 8

(b) Any outliers will be in regions: $x < 33 - 1.5 \times 8 = 21$

$x > 41 + 1.5 \times 8 = 53$

So there is only one outlier, 55.

(c)



Weight (kg)

(d) Normal distribution, as it is symmetric.

(e) They are not yet fully grown pigs.

9. (a)

0	0 3 5 6 7 9
1	0 2 3 4 6
2	2 4 5 6
3	3
4	5
5	
6	0 4
7	

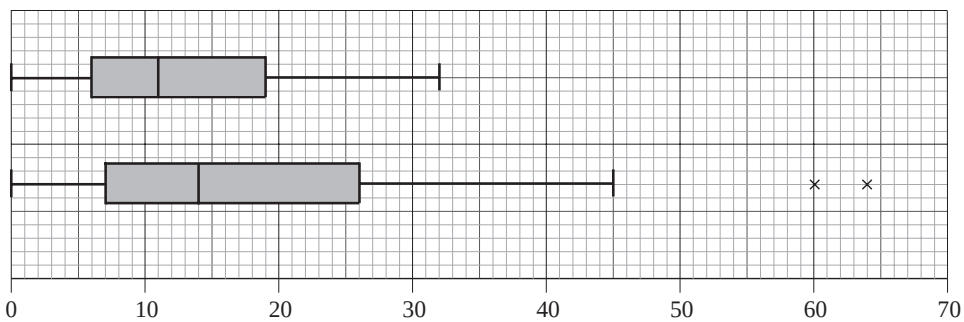
(b) Median = 14, LQ = 7, UQ = 26 (years)

(c) IQR = 26 – 7 = 19, so any outlier will be in regions: $x < 7 - 1.5 \times 19 = -21.5$

$x > 26 + 1.5 \times 19 = 54.5$

There are two outliers, 60 and 64.

(d)



Number of years

(e) Generally, Monarchs appear to reign longer than Popes, with a similar lower quartile, slightly greater median but significantly greater upper quartile, i.e. stretching to the right.

10. (a) (i) 9.5 (ii) $20 - 7 = 13$ (iii) About 8%
(b) (i) 8, 52, 67, 76, 83, 120 (ii) 13, 35.5

11. (a) 19 (b) 3 (c) (i) 46 (ii) 27.5, 58.5 (d) Graph (e) 32
(f)

Stem	*Leaf*
1	2 5 8
3	4
5	1 2 7

(g) 27%

12. (a) Graph (b) 7% (c) £14 000
(d) (i) £18 000 (ii) More variation for women (iii) Excludes extreme values
(e) (i) £13 300 (ii) No effect

13. (a) Positive (b) 33 minutes (c) 20 (d) 41 minutes

14. (a) Geometric mean = $(105 \times 140 \times 130)^{\frac{1}{3}} \approx 124$ (b) 24

20.4 Weighted Averages : Index Numbers

1. (a) It takes into account relative usage of commodities, so is a fairer representation.
(b) £500 (c) 117.25

2. (a) 133.33 (b) £14

3. (a) (i) 173.7 (ii) They have all increased by about 75%. (b) 176.0

4. (a) (i) 75 (ii) 150 (b) It remains the same at £400.

5. (a) 120.9 (b) 'All groups' index is less, but is based on many more categories.
(c) (i) 114.57
(ii) Reduced, since the 'mortgage' component has index well above 117.79.

20.5 Birth and Death Rates

1. (a) Bangladesh (b) 300 (c) (i) 306 (ii) $\frac{20}{1000} \times 15\,300$
(d) 50% of expected rate

2. (a) (i) 8 (ii) 6 (b) 17 535 (c) 111

3. (a) 15 (b) different distributions of ages, i.e. older population in Westport

(c) 19.96

(d) Westport's population distribution is not the same as the UK as a whole.

4. (a) 10 cm (b) 10, 3, 2, 5, 105

(c) Adamsville : 24.85; Beckbrough : 13.3

5. (a) $23 \div 12$, $23 \div 52$

(b) Heathfield : 83 / 6.9 / 1.6 Joseph : 234 / 19.5 / 4.5 Garath : 186 / 15.5 / 3.6

(c) 14

6. (a) P : 12.54, Q : 21.07

(b) Town P seems to be a more desirable place to live.

(c) 113, 108; good efficiency gains from 1986 to 1987.

20.6 Time Series Analysis : Moving Averages

1. (a) Graph (b) Trend line (c) 0.9

(d) The trend line is based on 1989 – 1994, so 1999 is too far away to predict.

2. (a) Trend line (b) 68

(c) The 20th week is too far away to make predictions.

3. (a) Underlying decreasing trend (b) Weeks 4 to 5

(c) (i) 4 (ii) 9 (d) 9.38, 9.20 (e) Graph

4. (a) 1992 (b) Closed for summer holidays (c) Graph

(d) (i)

Quarter	Turnover	Moving average
1	7500	
2	5500	5800
3	2000	
4	8200	5600
1	6700	5300
2	4300	5200
3	1600	5200
4	8200	4800
1	5100	4700
2	3900	4600
3	1200	4425
4	7500	

(ii) Graph

(iii) The underlying trend is downward.

5. (a) Graph (b) Spring/Summer time gives a boost to sales.

(c)

Year	Quarter	Sales £10 000's	Four-point moving average
1992	1	20	
	2	26	
	3	24	22
	4	18	23
			~~24~~
1993	1	24	24.75
	2	30	26
	3	27	26.5
	4	23	~~27.5~~
1994	1	26	28.5
	2	34	29
	3	31	30
	4	25	~~30.5~~
1995	1	30	31.5
	2	36	32.5
	3	35	
	4	29	

(d) Graph (e) Trend line (f) 33

6. (a) Graph
(b) 1st quarter is lowest; increase up to 3rd quarter; decrease in last quarter.
(c) If repeated in August and September, this would give 5.4 million passengers – you might though expect more passengers in July and less in September.

20.7 Correlation and Regression

1. B : positive correlation ; A : negative correlation ; C : no correlation

2. (a) E : 1600 cc ; F : 26 miles per gallon ; H : 2000 cc, 27 miles per gallon
(b) Diagram (c) G – it is well below a line of best fit.

3. (a) Diagram (b) (i) About 12.30 pm (ii) 38.5 °C (c) Aabout 6.35 pm

4. (b) 08.30, 21.15 (c) 01.00, 12.10

5. (a) and (c) Drawings (b) 27.5 kg (d) $3.4x + 1.6$ (e) 28.8 kg
(f) Small sample size

6. (a)

	A	*B*	*C*	*D*	*E*	*F*
Rank Time	3	5	2	4	1	6
Rank errors	4	2.5	5	2.5	6	1
d	1	2.5	3	1.5	5	5
d^2	1	6.25	9	2.25	25	25

(b) $\sum d^2 = 68.5$ and $r = -0.957$ (c) Quite fast and not too many errors.

7. (a) 0.006

   (b) Small but positive correlation, giving some support to the hypothesis.

   (c) Beer does not improve with age.

8. (a)

Club	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Position in league	1	3	6	2	7	8	5	4
Average attendance	34	12	18	32	15	25	27	19
Rank of attendance	1	8	6	2	7	4	3	5
Difference in ranks (d)	0	5	0	0	0	4	2	1
d^2	0	25	0	0	0	16	4	1

   (b) 0.45 (c) Some positive correlation

   (d) Very positive correlation

9. (a) 29% (b) Graph (c) 5.5 cm (d) Graph

   (e) (i) 8.2 cm (ii) 17.5 cm

   (f) Answer (i); lies in the range of chosen values for AB.

20.8 Distributions

1. (a) 122, 122, 122

   (b) (i) Identical (ii) On standardised scoring, all three are equivalent.

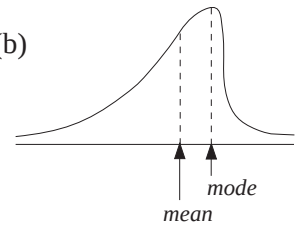
2. (a) Graph

   (b) Using mean 0 and s.d. 1, the standardised scores are
   History : 2 Physics : 1.5
   So this supports her claim.

3. (a) 5.5 (b) 0.76 (c) (i) 0.395 (ii) 5.7

   (d) (i) 3.0 (ii) Mean value increases, s.d. decreases

4. (a) Median (b)



5. (a) TTT, TTH, THT, HTT, THH, HTH, HHT, HHH

   (b) $\frac{1}{8}$ (c) $\frac{1}{2}$ (d) $\frac{1}{4}$

6. (a) 47 (b) Graph

7. Mean 120 cm; s.d. 3 cm

8. (a) Frequencies: 12, 40, 24, 15, 9

   (b) Graph

   (c) (i) 17.2 (ii) 8.5 lb

   (iii) Not likely; largest fish in frequency table $\leq$ 15 lb.

9. (a) (i) Too long (ii) Too expensive

(b) $417 \pm 3 \times 0.6 = 417 \pm 1.8$

$= 418.8 \text{ or } 415.2$

i.e. (415.2, 418.8)

(c) This ensures that more than 99.8% weigh more than 415 g.

(d) The process should be checked after 4 hours (sample 8) and $5\frac{1}{2}$ hours (sample 11), as both are outside the allowable limits.

10. (a) 1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1

(b) $\frac{6}{36} = \frac{1}{6}$

(c) $80 \times \frac{1}{6} \approx 13$

(d) $p = \frac{7}{9}$

$p(X = 2) = 3 \times \left(\frac{7}{9}\right)^2 \times \frac{2}{9} = \frac{98}{243}$

(e) The probability stays the same for each throw.

(f) She will get better with more throws.

11. (a)

4	3	8	2	7	2	6	8	9	3	2	1	5	0	0	8	5	2	2	6
G	G	R	B	R	B	Y	R	R	G	B	B	Y	P	P	R	Y	B	B	Y

(b) (i) $p(R) = \frac{5}{20} = \frac{1}{4}$

(ii) $100 \times \frac{1}{4} = 25$

(c) If the simulation is carried out 1000 times the probability might be slightly more than $\frac{1}{4}$ as the angle of Red on the spinner is greater than $90°$.