Revision Sheet 5 Worked Solutions

Q1.
(i) Any question that yields text data that can't be ordered:
e.g. What is your favourite subject at school?
(ii) Any question that yields numerical data that doesn't have certain fixed values
e.g. What is your height?

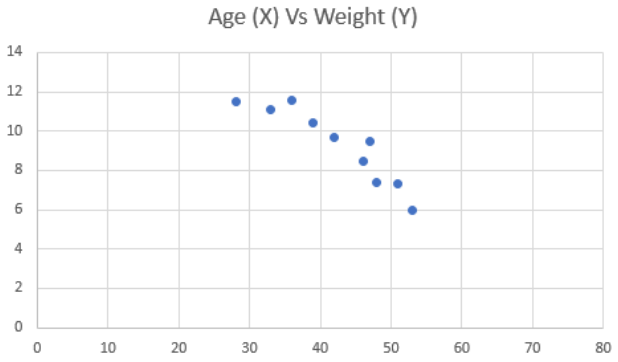
Q2. Any two of:
- Sample needs to be large enough
- Sample needs to be randomly selected
- Sample needs to be representative of the population

Q3.
Stratified Random Sample: The population is divided into two or more subgroups with similar characteristics and then a proportional sample is drawn from each subgroup. Would be a better choice than a simple random sample, if you wanted to see how opinions of particular sub-groups of the population vary.
Cluster Sample: The population is divided into clusters and then the clusters are selected randomly. This method might be cheaper than a simple random sample as the surveyor can just travel to the clusters to gather data rather than all over the country.

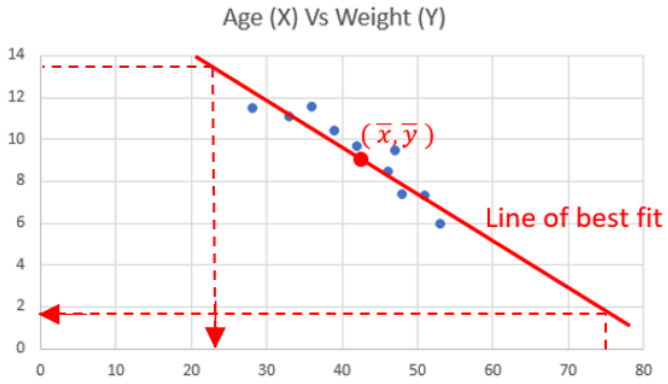
Q4.
(a)
- Use clear and simple language
- Avoid personal questions
- Start with simpler questions at the start
- Allow for all possible responses
- Be clear where answers should be recorded
- No leading questions

(b)
Advantages: Any one of: (i) Cheap (ii) responses can be anonymous => more honest answers
Disadvantages: Any one of: (i) Questions can't be explained to the respondent (ii) People don't always reply (iii) sample is biased as only people online surveyed

Q5.
(i) Mean Age
$= \frac{51 + 47 + 53 + 33 + 39 + 46 + 42 + 48 + 28 + 36}{10} = 42.3$
(ii) Mean Age
$= \frac{7.3 + 9.5 + 6 + 11.1 + 10.4 + 8.5 + 9.7 + 7.4 + 11.5 + 11.6}{10} = 9.3$
(iii)



(iv) Strong negative correlation
(v) (vi)



(vii) (a) From diagram above, expected weight of a 75-year old coin would be: 1.9g
(b) The expected age of a coin with a weight of 13.7g would be: 23 years
(viii) Not particularly reliable as our line of best fit is based on a very small amount of data (10 coins only) and the two values in the previous question are not even within the range of coin weights in the data. Would need a larger sample size if we wanted increased reliability.
(ix) r = -0.9252. This fits with the answer to part (iv) as it is a correlation coefficient showing strong negative correlation.

Q6.

(i) No. of values

= 24 + 16 + 42 + 40 + 12 = 134

To find the median, we add 1 to 134 and then divide by 2:

$= \frac{134+1}{2} = 67.5$

=> Median is average of 67th and 68th values

The first 24 values are in 0-20 range. The next 16 values are in 20-40 range and the next 42 values are in the 40-60 rang

=> the 67th and 68th values would be in the 40-60 range

=> Ans: 40 – 60 range

(ii) Mid-interval values are 10, 30, 50, 70, 90

=> Total paid out =

$(10 \times 24) + (30 \times 16) + (50 \times 42) + (70 \times 40) + (90 \times 12)$

= 6700 but these are €1000s of euro

=> amount paid out = €6,700,000

Q7. Spelling Test

i)

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

Key: 0|2 = 20 Key: 3|1 = 31

ii) Range A = Max – Min

= 42 – 15 = 27

Range B = Max – Min

= 47 – 14 = 33

=> A has greater range.

iii) 12 values => Median = $\frac{12+1}{2} = 6.5$

=> average of 6th & 7th values

Median A = $\frac{27+27}{2} = 27$

Median B = $\frac{22+26}{2} = 24$

Median of A slightly higher than B which suggests A performed slightly better.

iv) Mean A = $\frac{15+17+\cdots+42}{12} = 27.2$

Mean B = $\frac{14+18+\cdots+47}{12} = 28.2$

Mean of B is slightly higher than A which contradicts what the median told us. One person in group B got 47 out of 50, which compared to the highest value in group A, would pull the mean up a bit higher than A.

∴ IQR of A: Median = 6.5^{th} Value

Lower Quartile:

Median of lower 6 values

$= \frac{6+1}{2} = 3.5^{th}$ Value =

$= \frac{20+23}{2} = 21.5$

Upper Quartile:

Median of upper 6 values

$= 3.5^{th}$ Value $= \frac{31+33}{2} = 32$

$\Rightarrow$ IQR = 32 – 21.5 = 10.5

IQR of B:

$Q_1 = \frac{19+20}{2} = 19.5$

$Q_3 = \frac{35+39}{2} = 37$

$\Rightarrow$ IQR = 37 – 19.5

= 17.5

The higher IQR for B means the data in B is more widely spread about its centre than A.

vi) Easiest to use calculator as in handout but if doing by hand:

Mean of A = 27.2

$\Rightarrow$ Std dev

$= \sqrt{\frac{(31-27.2)^2+(17-27.2)^2+\cdots+(33-27.2)^2}{12}}$

= 7.7

Similarly, std dev for B is 10.6

As with the IQR, the std dev of B is higher than that of A, which means all the data is more spread out in B than in A.

Q8.

i) 13 students scored less than Sarah's 71

$\Rightarrow {}^{13}/_{20} = 65\%$ means she is in the 65th percentile $\boxed{P_{65}}$

ii) 20 values

$\Rightarrow 35\%$ of $(20+1) = 7.35$

$\Rightarrow$ Mean of 7th & 8th values

$= \frac{57+58}{2} = \boxed{57.5}$

iii) $P_{78} \Rightarrow 78\%$ of $(20+1)$

$= 16.38$

$\Rightarrow$ Mean of 16th & 17th values

$= \frac{76+79}{2} = \boxed{77.5}$

Q9.

i) By hand:

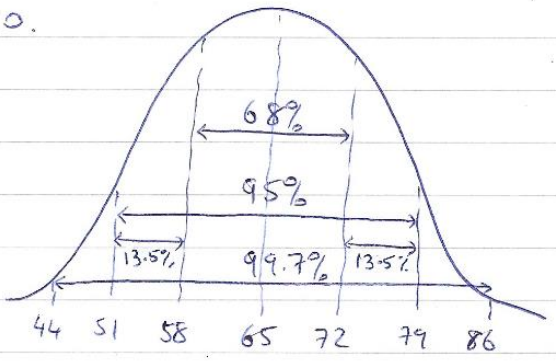
x Length	f Number	xf	$x-\bar{x}$ d	d^2	d^2f
5	4	20	-20	400	1600
15	16	240	-10	100	1600
25	20	500	0	0	20
35	12	420	10	100	1200
45	6	270	20	400	2400
	58	1450			6820

Mean $= \frac{1450}{58} = \boxed{25\text{cm}} = \bar{x}$

Std dev $= \sqrt{\frac{6820}{58}} = \boxed{10.8}$

ii) By calculator:
See handout from class for steps.

Q10.



i) Between 58 and 79 = 68% + 13.5%

$= \boxed{81.5\%}$

ii) Between 65 and 72 $= \frac{68\%}{2} = 34\%$

$\Rightarrow 34\%$ of $1000 = \boxed{340}$