Q1. **Statistics Revision Sheet Worked Solutions**

a) i) Discrete Numerical ii) Nominal Categorical iii) Continuous Numerical

iv) Ordinal Categorical

b) i) Face-to-face Interview/Telephone Interview/Online Questionnaire e.t.c.

ii) Advs: can explain the questions/can explain the questions/people more honest when not face-to-face…….

Disadvs: people might not be honest face-to-face/can be expensive/questions can't be explained…….

c) Assign every student in the school a number. Draw the numbers out of a hat or use the random number generator on your calculator to pick the numbers randomly.

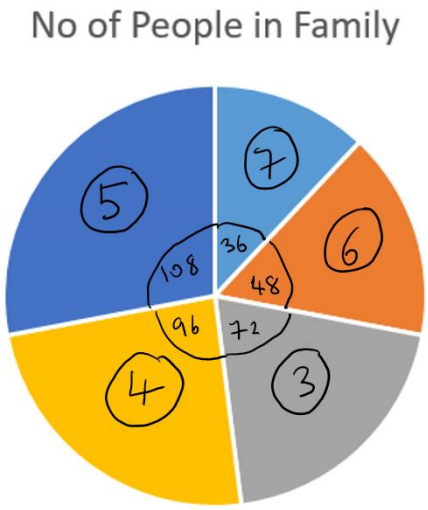
Q2.

a)

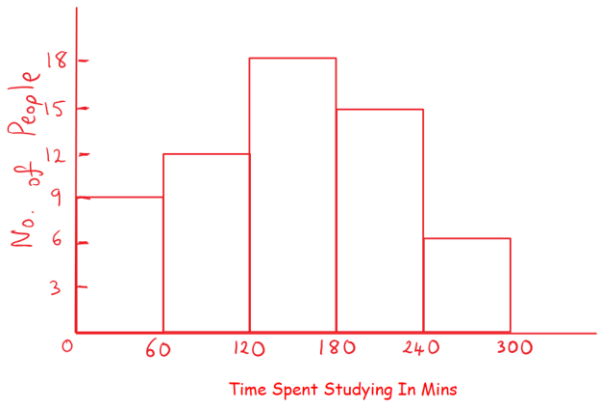
2	5
3	4, 6
4	1, 1, 4, 7
5	0, 2, 3, 5
6	5
7	1, 3, 6, 6, 6
8	2, 3

Key: 4|1 = 41%

b)

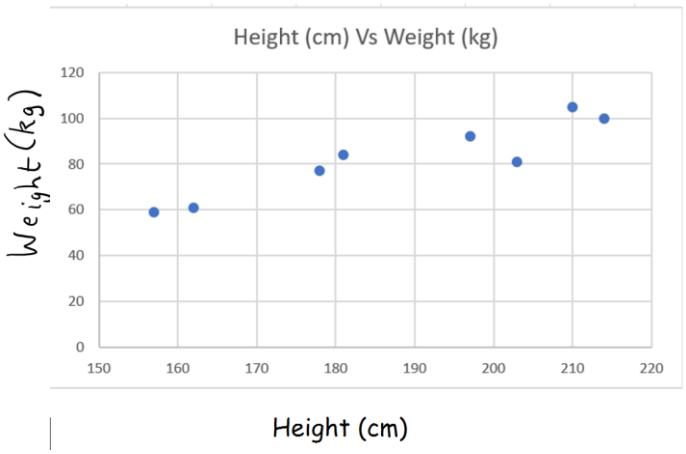


c) i)



ii) Shape of the distribution is symmetric (or normal)

d)



ii) Strong positive correlation – so generally, the taller you are, the heavier you are!

Q3.

a)

i) Mean $= \frac{2.4+3.1+2.5+3.1+1.8+3.4+2.7}{7}$

$= \frac{19}{7}$

$= \boxed{2.71}$

Median: Rerrange into order first

1.8, 2.4, 2.5, (2.7), 3.1, 3.1, 3.4

Median

Mode: One that appears most

$= \boxed{3.1}$

ii) Mean $= \frac{23+45+10+52+24+13+52+4}{8}$

$= \frac{223}{8}$

$= \boxed{27.88}$

Median: Rerrange into order first

4, 10, 13, (23, 24) 45, 52, 52

Median $\Rightarrow$ Median $= \frac{23+24}{2} = \boxed{23.5}$

Mode: One that appears most

$= \boxed{52}$

b) Mode or Median. There is a mode and neither are affected by the outlier 94.

Q4.

i)

Range	Mid Interval Values
0 – 60	$\frac{0+60}{2} = 30$
60 – 120	$\frac{60+120}{2} = 90$
120 – 180	$\frac{120+180}{2} = 150$
180 – 240	$\frac{180+240}{2} = 210$
240 – 300	$\frac{240+300}{2} = 270$

$\Rightarrow$ Mean $= \frac{(30 \times 9) + (90 \times 12) + (150 \times 18) + (210 \times 15) + (270 \times 6)}{9 + 12 + 18 + 15 + 6}$

$= \frac{270 + 1080 + 2700 + 3150 + 1620}{60}$

$= \frac{8820}{60} = \boxed{147 \text{ mins}}$

ii) The 6 students in the last category (240 – 300mins) are definitely included as they spend over 4 hours studying. The **maximum** number of students that could have spent over 3.5 hours would be if ALL 15 students in the second last category spent longer than 3.5 hours. So to answer the question, the maximum number of students in total that could have spent over 3.5 hours is 15 + 6 = 21 students.

Q5.

i) Range = Max Value − Min Value
= 19 − 6
= 13

ii) Need to rearrange data into order first:

6, 7, (8) LQ, 9, 9, 9, 11, 12, (15) UQ, 16, 19

– 11 values

Lower Quartile = Median of 1st 5 values
= $\frac{5+1}{2}$ = 3rd Value = 8

Upper Quartile = Median of Top 5 values
= 3rd value = 15

⇒ IQR = 15 − 8 = 7

Q6.

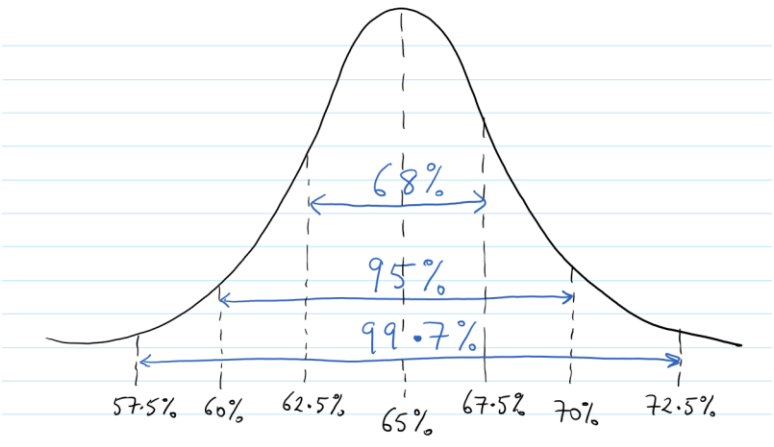
Need to calculate mean first:

$$\text{Mean} = \frac{(0\times 2)+(1\times 4)+(2\times 8)+(3\times 4)+(4\times 2)}{2+4+8+4+2} = 2$$

So, standard deviation will be:

$$\sqrt{\frac{(0-2)^2+(1-2)^2+(2-2)^2+(3-2)^2+(4-2)^2}{2+4+8+4+2}} = 1.09$$

Q7.



i) From diagram above, 68% of the class lie between 62.5% and 67.5%.

ii) From the diagram above, 95% of the class will be 2 standard deviations either side of the mean

iii) From diagram above, the percentage of students 1 standard deviation **above** the mean will be half of 68%, which is 34% => no. of students in the class = 34% of 30 = 10.2 = 10 students

Q8. $\hat{p} = \frac{12}{50} = 0.24$

$$E = \frac{1}{\sqrt{n}}$$
$$= \frac{1}{\sqrt{50}}$$
$$= 0.1414$$

$\Rightarrow \hat{p} - E < p < \hat{p} + E$

$0.24 - 0.1414 < p < 0.24 + 0.1414$

$0.0986 < p < 0.3814$
$9.8\% < p < 38.1\%$

→ Either answer is acceptable

Q9. $\hat{p} = \frac{35}{120} = 0.292$

$$E = \frac{1}{\sqrt{n}}$$
$$= \frac{1}{\sqrt{120}}$$
$$= 0.0913$$

$\Rightarrow \hat{p} - E < p < \hat{p} + E$

$0.292 - 0.0913 < p < 0.292 + 0.0913$

$0.2 < p < 0.38$
$20\% < p < 38\%$

Q10. $H_0: p = 0.43$

$H_1: p \neq 0.43$

$\hat{p} = \frac{202}{500} = 0.404$

$E = \frac{1}{\sqrt{500}} = 0.045$

$\Rightarrow \hat{p} - E < p < \hat{p} + E$

$0.404 - 0.045 < p < 0.404 + 0.045$

$0.359 < p < 0.449$

$\Rightarrow$ As 0.43 is in the range, we fail to reject H_0.

Q11. $H_0: p = 0.82$

$H_1: p \neq 0.82$

$\hat{p} = \frac{778}{1000} = 0.778$

$E = \frac{1}{\sqrt{1000}} = 0.032$

$\Rightarrow \hat{p} - E < p < \hat{p} + E$

$0.778 - 0.032 < p < 0.778 + 0.032$

$0.746 < p < 0.81$

As 0.82 is outside the range of our 95% Confidence Interval, we reject H_0 and accept H_1. So there isn't sufficient evidence to support the claim.

Past Exam Questions

Q12.

a)

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

Key: 1|9 = 19 yrs of age

b) $\% = \frac{14}{28} \times \frac{100}{1} = \boxed{50\%}$

Q13.

a) Range = max − min

i) $= 155 - 47$

$= \boxed{108\text{ mm}}$

ii) $\boxed{2003}$

iii) $\boxed{2002}$

b) 2006 : had the lowest rainfall and most amount of sun

c) 47, 72, 84, 94, 94, 101, 133, 134, 149, 155

10 values $\Rightarrow$ Median = $\frac{10+1}{2}$ = 5.5th Value

$\Rightarrow$ Median = $\frac{94+101}{2}$ = **97.5mm**

d) i) Mean = $\frac{169+124+180+173+173+239+159+168+228+205}{10}$

= $\frac{1818}{10}$ = **181.8 hrs**

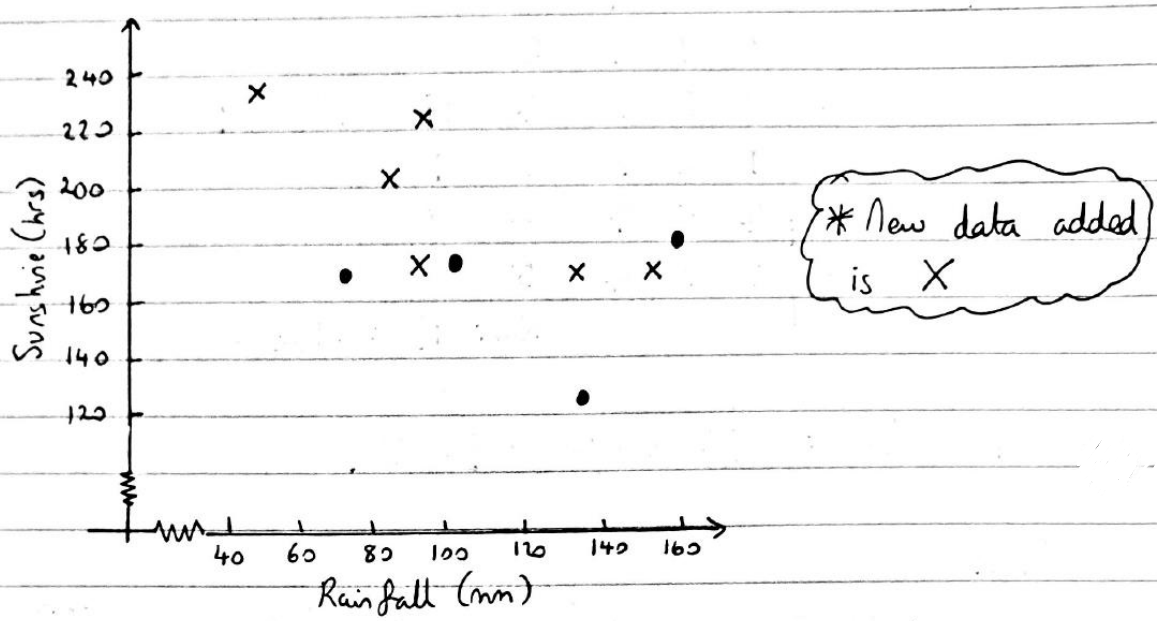
ii) 5% lower = 181.8 − (5% of 181.8) = 172.71

5% higher = 181.8 + (5% of 181.8) = 190.89

$\Rightarrow$ **2003, 2004, 2005**

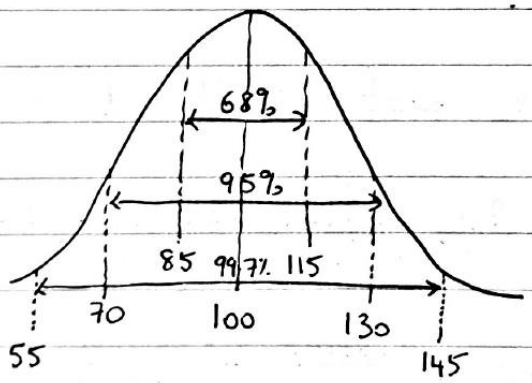
e) Using calculator : σ = 33.46 = **33.5**

f) i)



ii) −0.6 : Correlation is negative and reasonably strong

Q14. a) i)



ii) From diagram above : **95%**

iii) % between 85 + 115 is 68%

$\Rightarrow$ 68% of 1200 people = **816**

Q15. i) $E = \frac{1}{\sqrt{n}}$

$= \frac{1}{\sqrt{1200}}$

$= 0.0289$

$= 2.9\%$

$= \boxed{3\%}$

ii) $\hat{p} = \frac{578}{1200} = 0.48$

$\Rightarrow$ 95% Confidence Interval :

$\hat{p} - E < p < \hat{p} + E$

$0.48 - 0.03 < p < 0.48 + 0.03$

$\boxed{0.45 < p < 0.51}$

or $\boxed{45\% < p < 51\%}$

iii) $H_0 : p = 0.53$

$H_1 : p \neq 0.53$

From part (ii) Confidence Interval = $0.45 < p < 0.51$

As 0.53 is not in the range of our confidence interval we reject H_0 and accept H_1., so the claim they made was false.