Candidate No. A28437

Section A

N1

a) i. $\sum_{i=2}^{3} y_i^3 = y_2^3 + y_3^3 = 1 + (-8) = -7$

ii. $\sum_{i=1}^{3} \frac{x_i^2}{y_i} = \frac{x_1^2}{y_1} + \frac{x_2^2}{y_2} + \frac{x_3^2}{y_3} = \frac{16 \cdot 16}{64} + \frac{4 \cdot 4}{1} + \frac{2 \cdot 2}{-2} =$

$= 4 + 16 - 2 = 18$

iii. $|\sqrt{x_1}| + \sum_{i=2}^{3} y_i^{x_1} = 4 + 1^4 + (-2)^2 = 4 + 1 + 4 = 9$

b) i Categorical, nominal.
There is no method of measuring someones security number, and we can not put these numbers in the order, we know them only by their names.

ii Measurable. Because we can measure value of it.

iii Categorical, ordinal. Because final positions in the medals table of an Olympic games can be put in some sensible order. (Also categorical, because, there is no recognised method of measuring of their value)

c) i True, because it can be seen from the plot (by the median) (for example, if the median is closer to the top of the box, then distr. is skewed left)

ii False, because A and B can be independent

iii False, because $E[X] = 4$

iv True. $\frac{1}{6} \cdot 120 = \frac{120}{6} = 20$ - integer

v. False. Due to the slope coefficient formula

Candidate No. A28437

N1

d) Characteristics of experimental study:
- can be repeated
- can be controlled

example:

e) i. $E(x) = \sum_{i=1}^{N} p_i x_i = 2\cdot 0{,}25 + 4\cdot 0{,}25 + 6\cdot 0{,}40 + 8\cdot 0{,}10 = 4{,}7$

ii. $\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} p_i (x_i - \mu)^2} =$

$= \sqrt{\sum_{i=1}^{4} p_i (x_i - 4.7)^2} = \sqrt{0{,}25\cdot(2-4.7)^2 + 0{,}25 + (4-4.7)^2 + 0{,}40(6-4.7)^2 + 0.10(8-4.7)^2} = \sqrt{3{,}71} \approx 1{,}9261$

iii. $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)} =$

$= 0{,}8667$

iv. X does not have a uniform distribution

Candidate No. A28437

N1

f)

Product	A	B	C	D	E	F	G	H
Shopper 1	6	5	8	2	3	7	1	4
Shopper 2	7	8	6	1	2	5	4	3
Rank of Shopper 1	6	5	8	2	3	7	1	4
Rank of Shopper 2	7	8	6	1	2	5	4	3
difference in ~~values~~ ranks, d_i	-1	-3	-2	1	1	2	-3	1
d_i^2	1	9	4	1	1	4	9	1

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(1+9+4+1+1+4+9+1)}{(8^2-1)8} =$$

$$= 1 - \frac{180}{504} = 1 - \frac{5}{14} = \frac{9}{14} = 0.6428$$

There is quite strong correlation. There is a positive connection in terms of ranks.

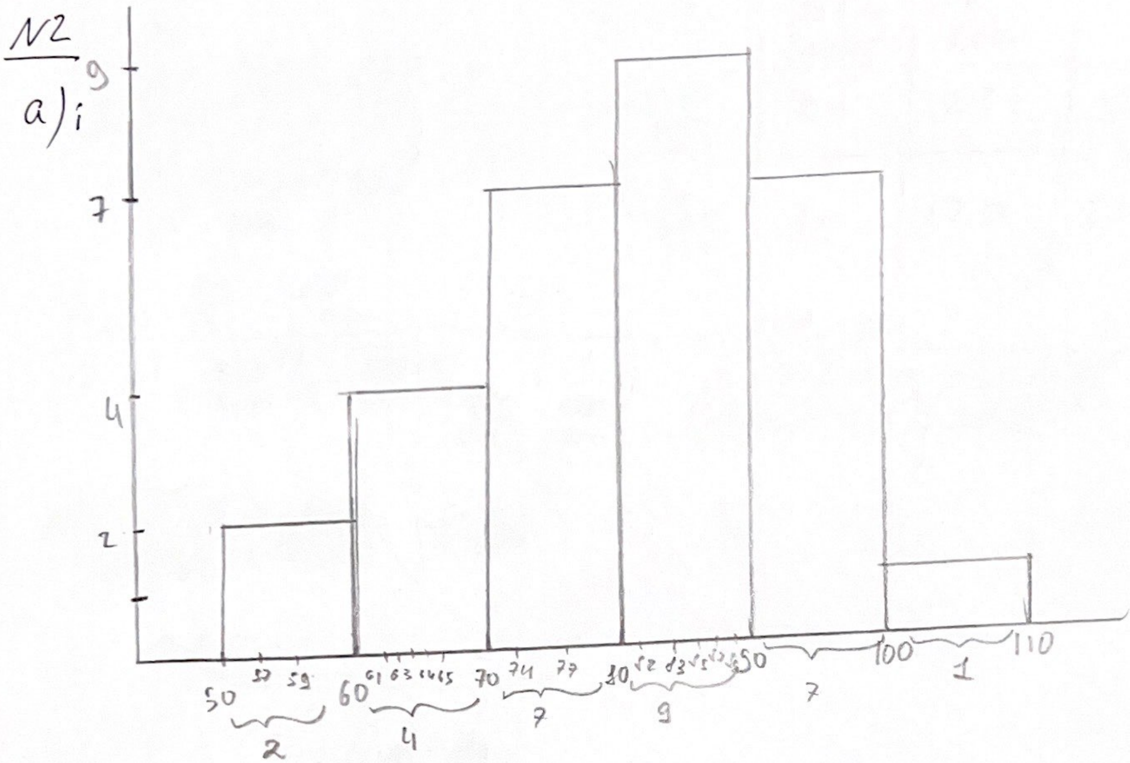
g) i $n=240$

Candidate No. A28437.

Section B

N2

a) i



ii Mean

$\frac{1}{30}(57+59+61+63+64+65+\ldots+99+101) = \frac{242}{3} \approx 80.67$

Modal group [80; 90], as we do not have 80 and 90 in our sample, we can say that modal group is [81, 89]

iii Median is $\hat{med} = 82$ (as $\frac{82+82}{2} = 82$)

lower quartile is

$Q_1 = X_{(8)} = 74$

iv ~~The distribution of the data appears to be a little big right skewed as the me~~

The distribution of the data appears to be almost symmetrical. This is also supported by the fact that median is almost equal to the mean value. But we can also say that our data is a little bit left-skewed

Candidate No. A28437

№2

б/ $H_0: p_1 = p_2$

$H_A: p_1 \neq p_2$

	not in favour	in favour	
40 years old +	104	221	325
40 years old −	80	120	200

$$T = \frac{\frac{221}{325} - \frac{120}{200}}{\sqrt{\frac{184}{500}\left(1 - \frac{184}{500}\right)\left(\frac{1}{325} + \frac{1}{200}\right)}}$$

$$= \frac{0{,}08}{\sqrt{\frac{184}{500}\left(1 - \frac{184}{500}\right)\left(\frac{1}{325} + \frac{1}{200}\right)}} = 1{,}8$$

$T|_{H_0} \sim N(0;1)$

Let $\alpha = 0{,}05$

$z_{0{,}05} = -1{,}65$

−1,65 1,65 1,8

Then we have to reject H_0 in favour of H_A at $\alpha = 0{,}05$

Let $\alpha = 0{,}01$

$z_{0{,}01} = -1{,}33$

−1,33 1,33 1,8

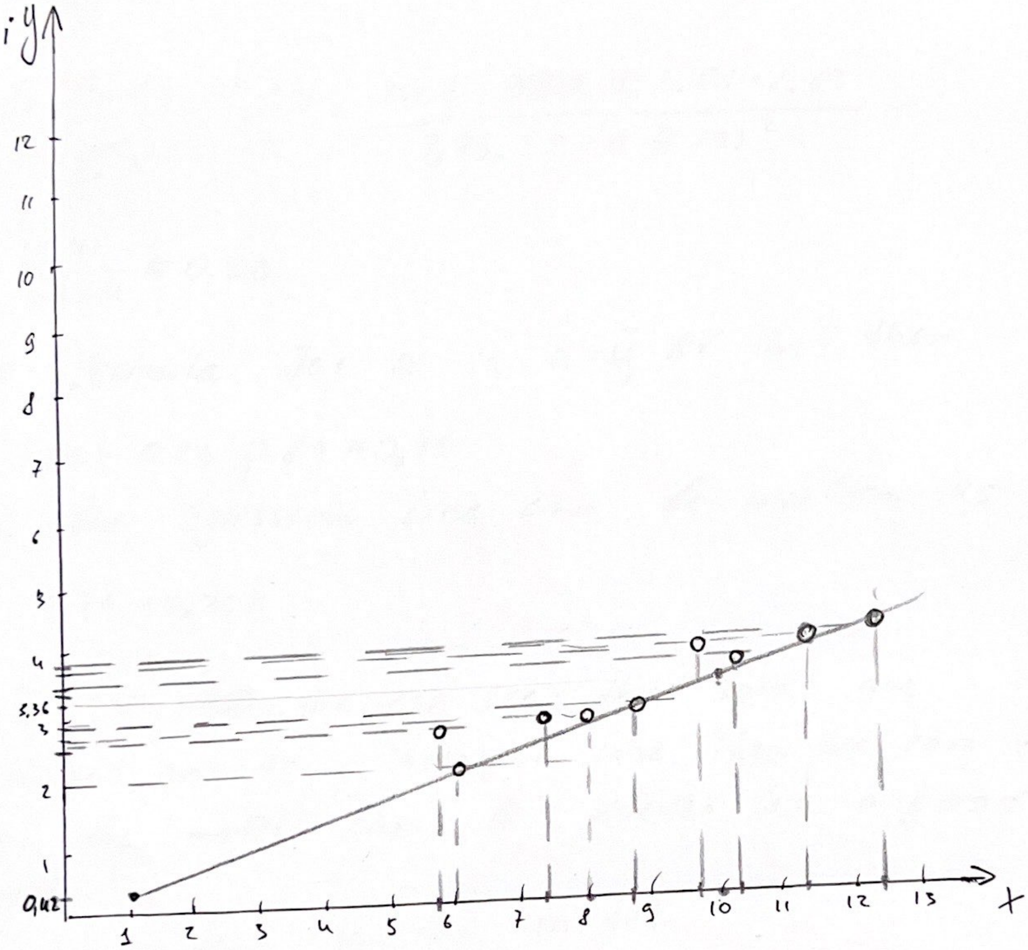
Then we have to reject H_0 in favour of H_A at $\alpha = 0{,}01$

~~Then~~

Candidate No. A28437

N3

a) i



ii Sample correlation coefficient

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y^2 - n\bar{y}^2\right)}} =$$

$$= \frac{254.6 - 9 \cdot 3.06 \cdot 8.83}{\sqrt{(745.37 - 9 \cdot (8.83)^2)(87.82 - 9 \cdot (3.06)^2)}} =$$

$$= \frac{254.6 - 9 \cdot 3.06 \cdot 8.83}{\sqrt{43,64 \cdot 3,54}} = \frac{254.6 - 9 \cdot 3.06 \cdot 8.83}{12,43} =$$

$$= \frac{11,42}{12,43} \approx 0,92$$

$$\bar{x} = (6 + 9.7 + 8 + 11.4 + 8.7 + 5.7 + 10.3 + 7.3 + 12.4)\frac{1}{9} =$$

$$= \frac{159}{18} \approx 8,83$$

$$\bar{y} = (2 + 3.7 + 2.7 + 3.7 + 2.9 + 2.6 + 3.5 + 2.7 + 3.8)\frac{1}{9} =$$

$$= \frac{46}{15} \approx 3,06$$

Candidate No. A28437

Due to the fact that r is closer to 1, we can say that it is strong linear positive association

iii. $b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - nx^2} = \frac{254.6 - \cancel{\ } 9 \cdot 3.06 \cdot 8.83}{745.37 - 9 \cdot (8.83)^2} =$

$= \frac{11,42}{43.64} \approx 0.26$

The formula for a is $a = \bar{y} - b\bar{x}$ and then

$a = 3.06 - 0.26 \cdot 8.83 \approx 0,76$

Then the regression line can be written as

$\hat{y} = 0,76 + 0,26x$.

iv. In this case we can see that point are scattered around a straight line. Then we can conclude that the model can be rejected as a good model

Then according to the model

$0,76 + 0,26 \cdot 8 = \underline{2,84}$

A person, who follows this program will loose aproximately 2,84 kg.

I would trust this value, because, as I already said, the model seems to be good and because results, that are given in the terms of the task ~~(8;2.7)~~ say that someone, who was following the diet for 8 weeks had lost 2.7 kg.