HW 2 Part 1

1)

A.

Mean

Bin 1: 15, 15, 16, 16, 19 → 16.2, 16.2, 16.2, 16.2, 16.2

Bin 2: 20, 20, 21, 22, 22 → 21, 21, 21, 21, 21

Bin 3: 25, 25, 30, 33, 33 → 29.2, 29.2, 29.2, 29.2, 29.2

Bin 4: 36, 40, 46, 52, 70 → 48.8, 48.8, 48.8, 48.8, 48.8

Bin 1: 16.2, 16.2, 16.2, 16.2, 16.2

Bin 2: 21, 21, 21, 21, 21

Bin 3: 29.2, 29.2, 29.2, 29.2, 29.2

Bin 4: 48.8, 48.8, 48.8, 48.8, 48.8

B.

Bin 1: 18-26 → 18, 18, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25

Bin 2: 26-37 → 30, 33, 33, 36

Bin 3: 37 - 48 → 40, 46

Bin 4: 48 - 59 → 52

Bin 5: 59 - 70 → 70

now appb boudaries

Bin 1: 15, 15, 15, 15, 15, 15, 15, 25, 25, 25, 25, 25

Bin 2: 30, 33, 33, 36

Bin 3: 40, 46

Bin 4: 52

Bin 5: 70

2)

$$\chi^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

	Group 1	Group 2	total
Student	40 45	20 15	60
Employed	95 90	25 30	120
Self-employed	15 15	5 5	20
total	150	50	200

Expected Values in blue

e(x) = # col $\uparrow_r$ # col ~~row~~ row / total

$$\chi^2 = \frac{(40-45)^2}{45} + \frac{(20-15)^2}{15} + \frac{(95-90)^2}{90} + \frac{(25-30)^2}{30} + \cancel{\frac{(15-15)^2}{15}} + \cancel{\frac{(5-5)^2}{5}} = 3.333$$

$$\boxed{\chi^2 = 3.33}$$

3)

A.

age	23	23	27	27	39	41	47	49
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2

$$\bar{A} = \frac{23+23+27+27+39+41+47+49}{8} = 34.5$$

$$\bar{B} = \frac{9.5+26.5+7.8+17.8+31.4+25.9+27.4+27.2}{8} = 21.6875$$

$$\sigma_A^2 = 1/8\big((23-34.5)^2 + (23-34.5)^2 + (27-34.5)^2 + (27-34.5)^2 + (39-34.5)^2 + (41-34.5)^2 + (47-34.5)^2\big) = 100.75$$

$$\sigma_B^2 = 1/8\big((9.5-21.6875)^2 + (26.5-21.6875)^2 + (7.8-21.6875)^2 + (17.8-21.6875)^2 + (31.4-21.6875)^2 + (25.9-21.6875)^2 (27.4-21.6875)^2 + (27.2-21.6075)^2 = 69.34609$$

$$\sigma_A = \sqrt{100.75} = 10.04$$

$$\sigma_b = \sqrt{69.346} = 8.33$$

$$r_{A,B} = \frac{23\cdot 9.5 + 23\cdot 26.5 + 27\cdot 7.8 + 27\cdot 17.8 + 39\cdot 31.4 + 41\cdot 25.9 + 47\cdot 27.4 + 49\cdot 27.2}{8\cdot 10.04\cdot 8.33} = \boxed{0.659}$$

So A,B positivly corellated

B.

$$r_{A,B} = \frac{23 \cdot 9.5 + 23 \cdot 26.5 + 27 \cdot 7.8 + 27 \cdot 17.8 + 39 \cdot 31.4 + 41 \cdot 25.9 + 47 \cdot 22.5}{8}$$

$$\underline{49 \cdot 27.2} - 34.5 \cdot 21.6875 = \boxed{55.06875}$$

$\boxed{\text{So A and b move in the same direction}}$

4)

A. $$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

$$200 \rightarrow \frac{200-200}{1005-200} = 0$$

$$300 \rightarrow \frac{300-200}{1005-200} = 0.124$$

$$400 \rightarrow \frac{400-200}{1008-200} = 0.248$$

$$600 \rightarrow \frac{600-200}{1005-200} = 0.497$$

$$1005 \rightarrow \frac{1005-200}{1005-200} = 1$$

$$\boxed{\text{Normalized Data} = (0, 0.111, 0.222. \ 0.444, 1)}$$

B.

$$\mu = \frac{200 + 300 + 400 + 600 + 1005}{5} = 501$$

$$\sigma = \sqrt{\frac{(200-501)^2 + (300-501)^2 + (400-501)^2 + (600-501)^2 + (1005-501)^2}{5}} = 296.67$$

$$X_{norm} = \frac{X - \mu}{\sigma}$$

$$200 \rightarrow \frac{200-501}{296.67} \rightarrow -1.044$$

$$300 \rightarrow \frac{300-501}{296.57} \rightarrow -0.674$$

$$400 \rightarrow \frac{400-501}{296.57} \rightarrow -0.409$$

$$600 \rightarrow \frac{600-501}{296.57} \rightarrow 0.303$$

$$1005 \rightarrow \frac{1005-501}{296.57} \rightarrow 1.819$$

$$\boxed{\text{Norminalized Data} = (-1.044, -0.674, -0.404, 0.303, 1.819)}$$

c. scalle 10^4

$200 \to 200/10^4 \to 0.02$

$300 \to 300/10^4 \to 0.03$

$400 \to 400/10^4 \to 0.04$

$600 \to 600/10^4 \to 0.06$

$1005 \to 1005/10^4 \to 0.1005$

$\boxed{\text{Norminalized data} = (0.02, 0.03, 0.04, 0.06, 0.1005)}$

5a)

Support $\{1,2\} \to \{3\} = 3/5 = 0.6$

Confidence $\{1,2\} \to \{3\} = 3/4 = 0.75$

Support $\{5\} \to \{1,2\} = 3/5 = 0.6$

Confidence $\{5\} \to \{1,2\} = 3/3 = 1$

5b)

Join $\{1,2\}, \{1,3) \to (1,2,3)$

Join $\{1,2\}, \{1,5\} \to (1,2,5)$

Join $\{1,2\}, \{2,3) \to (1,2,3)$

Join $\{1,2\}, \{2,5\} \to (1,2,5)$

Join $\{1,3\}, \{1,5\} \to (1,3,5)$

Join $\{1,3\}, \{1,5\} \to (1,2,3)$

Join $\{1,3\}, \{1,5\} \to (1,3,5)$

Join $\{1,3\}, \{2,3\} \to (1,2,3)$

join $\{1,5\}, \{2,5\} \to (1,2,5)$

join $\{2,3\}, \{2,5\} \to (2,3,5)$

Remove duplicates

$$\boxed{C_3 = \{(1,2,3), (1,2,5), (1,3,5), (2,3,5)}$$

5c) Possible Association rules are

$\{1,2\} \to 3$, $\{2,3\} \to 1$, $\{1,3\} \to 2$,

$1 \to (2,3)$, $\cancel{2 \to (1,3)}$, $3 \to (1,2)$

Not confident enough

confidence $\{1,2\} \to 3$ is $\dfrac{(1,2,3)\text{ count}}{(1,2)\text{ count}}$

$= 3/4 = .75$

thus confident at 75%

confidence $\{2,3\} \to 1$ is $\dfrac{(1,2,3)\text{ count}}{(2,3)\text{ count}}$

$= 3/3 = 1$

thus confident at 100%

confidence $\{1,3\} \to 2$ is $\dfrac{(1,2,3)\text{ count}}{(1,3)\text{ count}}$

$= 3/4 = .75$

thus confident at 75%

confidence $1 \to \{2,3\}$ is $\dfrac{(1,2,3)\text{ count}}{1\text{ count}}$

$= 3/5 = .60$

thus confident at 60%
not enough

Not confident enough

confidence $2 \to \{1,3\}$ is $\dfrac{(1,2,3)\text{ count}}{2\text{ count}}$

$= 3/4 = .75$

thus confident at 75%

confidence $3 \to \{1,2\}$ is $\dfrac{(1,2,3)\text{ count}}{3\text{ count}}$

$= 3/4 = .75$

thus confident at 75%