1) 15, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 30, 33, 33, 36, 40, 46, 52, 70

a) bin means, depth = 5

15, 15, 16, 16, 19 | 20, 20, 21, 22, 22 | 25, 25, 30, 33, 33, 36, 40, 46, 52, 70

$\frac{2(15)+16+16+19}{5} = 16.2$ x5 , $\frac{2(20)+21+2(22)}{5} = 21$ x5 , $\frac{2(25)+30+2(33)}{5} = 29.2$ x5 , $\frac{36+40+46+52+70}{5} = 48.8$ x5

b) bin boundaries, bin width = 11

$\frac{70-15}{11} = 5 =$ # of bins

i.e. each bin has 4 elements.

15, 15, 16, 16 | 19, 20, 20, 21 | 22, 22, 25, 25 | 30, 33, 33, 36 | 40, 46, 52, 70

✓ | ↓ 19, 19, 21, 21 | ✓ | ↓ 30, 30, 36, 36 | ↓ 40, 40, 40, 70

⇒ 15, 15, 16, 16, 19, 19, 21, 21, 22, 22, 25, 25, 30, 30, 36, 36, 40, 40, 40, 70

2)

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

$$\chi^2 = \frac{(40-45)^2}{45} + \frac{(20-15)^2}{15} + \frac{(95-90)^2}{90} + \frac{(25-30)^2}{30} + 0 + 0$$

$$= \frac{25}{45} + \frac{25}{15} + \frac{25}{90} + \frac{25}{30}$$

$$= 10/3$$

3)

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

a) Calculate r

let A = age, B = % fat, n = 8

$\overline{A} = 34.5 \longrightarrow \sigma_A = 10.04$

$\overline{B} = 21.6875 \longrightarrow \sigma_B = 8.33$

A_i	$A_i - \overline{A}$	B_i	$B_i - \overline{B}$	$(A_i - \overline{A})(B_i - \overline{B})$
23	-11.5	9.5	-12.1875	140.15625
23	-11.5	26.5	4.8125	-55.34375
27	-7.5	7.8	-13.8875	104.15625
27	-7.5	17.8	-3.8875	29.15625
39	4.5	31.4	9.7125	43.70625
41	6.5	25.9	4.2125	27.38125
47	12.5	27.4	5.7125	71.40625
49	14.5	27.2	5.5125	79.93125

$$\frac{440.55}{8}$$

$$= 55.06875$$

$$= COV(A,B)$$

$$r_{A,B} = \frac{COV(A,B)}{\sigma_A \sigma_B} = \frac{55.06875}{10.04(8.33)}$$

$$= 0.6585$$

↑ positive correlation

b) Calculate covariance

From before, $COV(A,B) = 55.06875$

4) Use the following methods to normalize the following dataset:

200, 300, 400, 600, 1000

a) min-max : min = 0, max = 1

$max - min = 1000 - 200 = 800$

$0, \frac{1}{8}, \frac{1}{2}, \frac{3}{4}, \frac{5}{4}$

b) z-score

$$\mu = \frac{1500}{5} = 300$$

$$\sigma = \sqrt{\frac{10,000 + 0 + 10,000 + 90,000 + 490,000}{5}}$$

$$= \sqrt{\frac{600,000}{5}} = \sqrt{120,000}$$
$$= 200\sqrt{3}$$

$$\frac{-100}{200\sqrt{3}}, 0, \frac{100}{200\sqrt{3}}, \frac{300}{200\sqrt{3}}, \frac{700}{200\sqrt{3}}$$

$$\frac{-1}{2\sqrt{3}}, 0, \frac{1}{2\sqrt{3}}, \frac{3}{2\sqrt{3}}, \frac{7}{2\sqrt{3}}$$

---

c) decimal scaling

$\Rightarrow .02, .03, .04, .06, .1$