

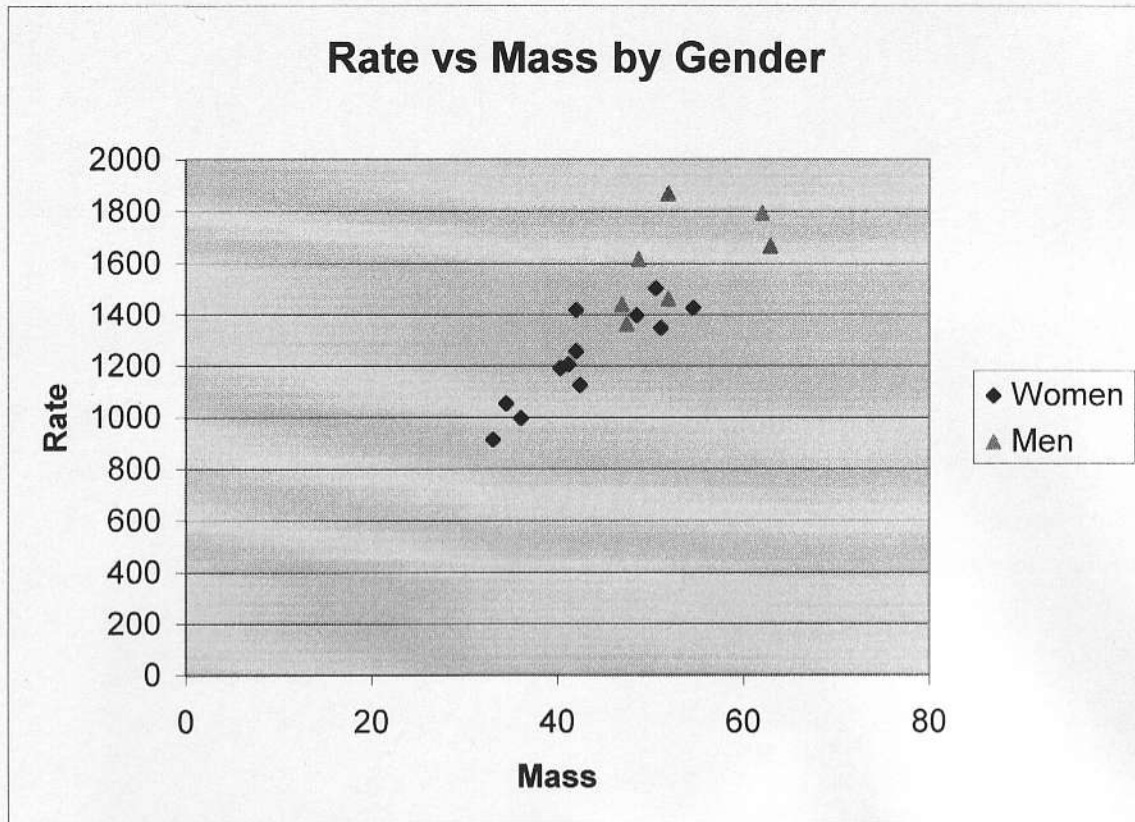
Homework #2 Solutions

2.11, 2.18, 2.20, 2.21, 2.31

Problem 2.11

- (a) see attached plot
- (b) The association is positive, roughly linear and moderate to strong in nature. The relationship seems roughly the same for men and women but the men are generally higher in both mass and rate.

Scatterplot for 2.11(a)



Problem 2.18

(a) See attached plot

$$\begin{aligned}\text{mean Yellow} &= \frac{45 + 59 + 48 + 46 + 38 + 47}{6} \\ &= 47.16\end{aligned}$$

$$\begin{aligned}\text{Mean white} &= \frac{21 + 12 + 14 + 17 + 13 + 17}{6} \\ &= 15.66\end{aligned}$$

$$\begin{aligned}\text{Mean Green} &= \frac{37 + 32 + 15 + 25 + 39 + 41}{6} \\ &= 31.5\end{aligned}$$

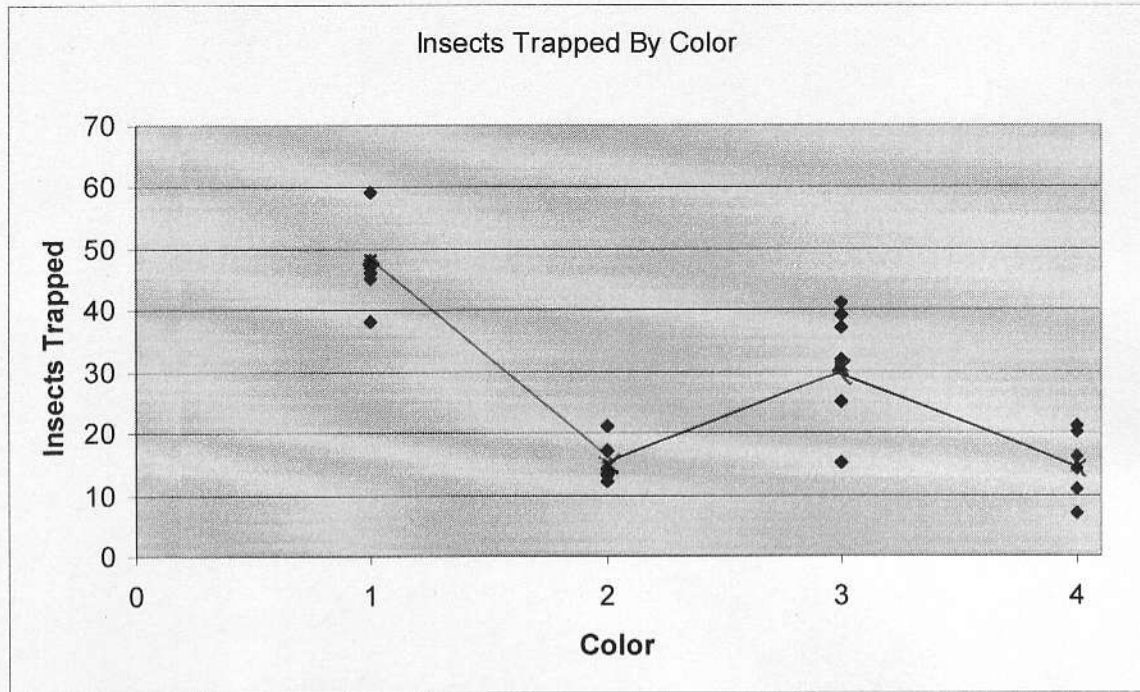
$$\begin{aligned}\text{mean blue} &= \frac{16 + 11 + 20 + 21 + 14 + 7}{6} \\ &= 14.83\end{aligned}$$

(b) Lemon yellow is most attractive
Green is also attractive.

white and blue are not very attractive

(c) No color is not a numerical variable
it is categorical - Also colors are not ordered.

Figure for problem 2.19



Problem 2.20

- (a) See the plot for exercise 2.11
- (b) To find r for women alone and for men alone use only observation of women and then only of men
Note X is Mass, Y is rate
First for women

$$n_{\text{women}} = 12$$

$$\begin{aligned}\sum_{i=1}^{12} X_i &= 36.1 + 54.6 + 48.5 + 42.0 + 50.6 + 42.0 \\ &\quad + 40.3 + 33.1 + 42.4 + 34.5 + 51.1 + 41.2 \\ &= 516.4\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{12} Y_i &= 995 + 1425 + 1396 + 1419 + 1502 + 1256 + 1189 + 913 \\ &\quad + 1124 + 1052 + 1347 + 1204 \\ &= 14821\end{aligned}$$

$$\bar{X} = 516.4/12 = 43.03$$

$$\bar{Y} = 14821/12 = 1235.08$$

$$\sum_{i=1}^{12} x_i^2 = 36.1^2 + 54.6^2 + \dots + 41.2^2$$

$$= 22741.34$$

$$\sum_{i=1}^{12} y_i^2 = 995^2 + 1425^2 + \dots + 1204^2$$

$$= 18695125$$

$$s_x = \sqrt{\frac{22741.34 - 12(43.03)^2}{12-1}}$$

$$= 6.86$$

$$s_y = \sqrt{\frac{18695125 - 12(1235.08)^2}{12-1}}$$

$$= 188.28$$

$$\sum_{i=1}^{12} x_i y_i = (36.1)(995) + (54.6)(1425) + \dots + (41.2)(1204)$$

$$= 650264.8$$

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

$$= \frac{1}{11} \frac{1}{6.86} \frac{1}{188.28} (650264.8 - 12(43.03)(1235.08))$$

$$= 0.88 \text{ (2dp)}$$

For men

$$n=7$$

$$\begin{aligned}\sum_{i=1}^7 x_i &= 62 + 62.9 + \dots + 46.9 \\ &= 371.7\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^7 y_i &= 1792 + \dots + 1439 \\ &= 11200\end{aligned}$$

$$\bar{x} = 371.7/7 = 53.1$$

$$\bar{y} = 11200/7 = 1600$$

$$\begin{aligned}\sum_{i=1}^7 x_i^2 &= 62^2 + 62.9^2 + \dots + 46.9^2 \\ &= 20005.69\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^7 y_i^2 &= 1792^2 + 1666^2 + \dots + 1439^2 \\ &= 18134870\end{aligned}$$

$$s_x = \sqrt{\frac{20005.69 - 7(53.1)^2}{7-1}} = 6.6885$$

$$s_y = \sqrt{\frac{18134870 - 7(1600)^2}{7-1}} = 189.2397$$

$$\sum_{i=1}^7 X_i Y_i = (62.0)(1792) + (62.1)(1666) \\ + \dots + (46.9)(1439) \\ = 599216.4$$

$$r = \frac{1}{n-1} \left(\frac{1}{s_x} \right) \left(\frac{1}{s_y} \right) \left(\sum_{i=1}^n X_i Y_i - n \bar{x} \bar{y} \right) \\ = \frac{1}{6} \left(\frac{1}{6.63} \right) \left(\frac{1}{189.23} \right) (599216.4 - 6(53.1)(1600)) \\ = 0.59 \text{ (2dp)}$$

$$(c) \quad \bar{X}_{\text{women}} = 43.03$$

$$\bar{X}_{\text{men}} = 53.1$$

Correlation is not affected by changes in location so it is not affected by the change in mean body mass.

(d) the correlations would not change since it is unitless

Problem 2.21

(a) see attached plot. From the plot we see that higher mens heights are more likely to be found with higher women heights so we would expect the correlation to be positive.

Since the points do not fall close to or along a line we would not expect it to be close to 1.

(b) To Find the correlation we need
~~first~~

1. $n = 6$

2.
$$\sum_{i=1}^6 x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$
$$= 66 + 64 + 66 + 65 + 70 + 65$$
$$= 396$$

$$\sum_{i=1}^6 y_i = y_1 + y_2 + y_3 + y_4 + y_5 + y_6$$
$$= 72 + 68 + 70 + 68 + 71 + 65$$
$$= 414$$

$$3. \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{396}{6} = 66$$

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{414}{6} = 69$$

$$4. \sum_{i=1}^6 X_i^2 = X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2$$

$$= 66^2 + 64^2 + 66^2 + 65^2 + 70^2 + 65^2$$

$$= 26158$$

$$\sum_{i=1}^6 Y_i^2 = Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 + Y_6^2$$

$$= 72^2 + 68^2 + 70^2 + 68^2 + 71^2 + 65^2$$

$$= 28598$$

$$5. S_X = \sqrt{\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{26158 - 6(66)^2}{6-1}}$$

$$= \sqrt{\frac{26158 - 26136}{5}}$$

$$= \sqrt{\frac{22}{5}} = 2.0976$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}{n-1}} = \sqrt{\frac{28598 - 6(69)^2}{6-1}}$$

$$= \sqrt{\frac{28598 - 28566}{5}}$$

$$= \sqrt{\frac{32}{5}} = 2.5298 \text{ (4dp)}$$

$$\begin{aligned} 6. \sum_{i=1}^6 x_i y_i &= x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 + x_5 y_5 + x_6 y_6 \\ &= (66)(72) + (64)(68) + (66)(70) + (65)(68) + (70)(71) + (65)(65) \\ &= 27339 \end{aligned}$$

$$7. r = \frac{1}{n-1} \left(\frac{1}{s_x} \right) \left(\frac{1}{s_y} \right) \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)$$

$\uparrow$ from 1 $\uparrow$ from 5 $\uparrow$ from 6 $\uparrow$ from 3

$$r = \frac{1}{6-1} \left(\frac{1}{2.0976} \right) \left(\frac{1}{2.5298} \right) (27339 - 6(66)(69))$$

$$= \frac{1}{5} \left(\frac{1}{2.0976} \right) \left(\frac{1}{2.5298} \right) (15)$$

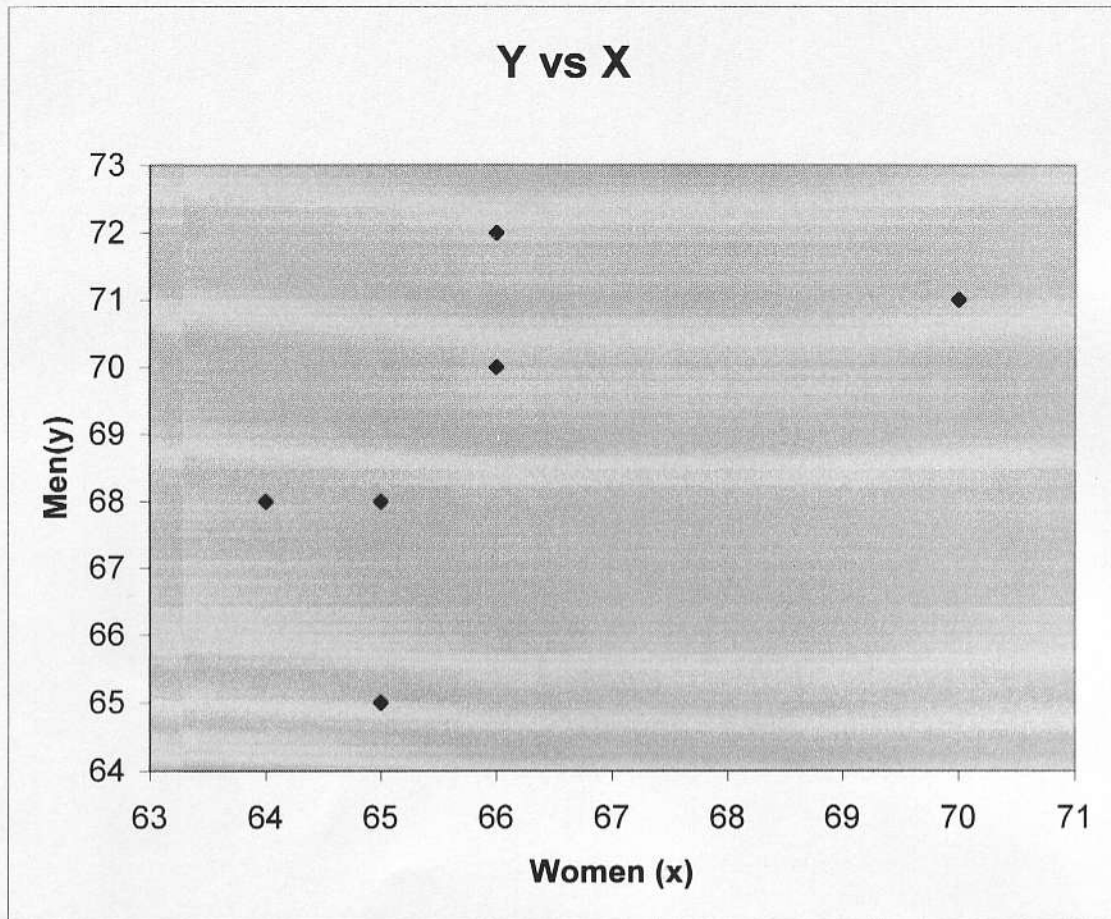
$$= 0.57 \text{ (2dp)}$$

(c) The correlation would remain unchanged. It does not tell us whether the men are taller.

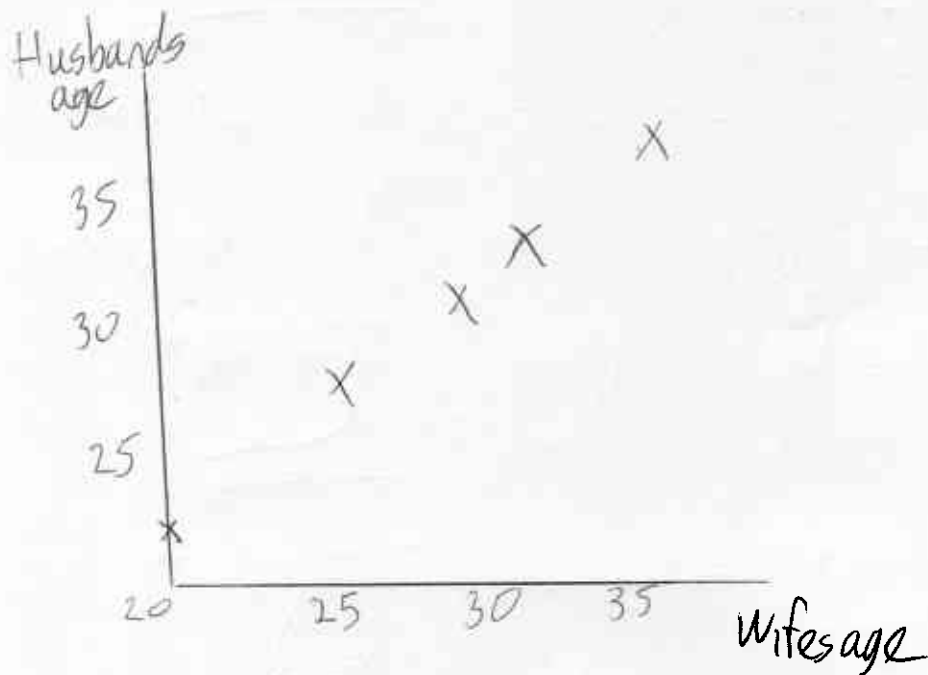
(d) Correlation is unitless so it would not change

(e) The correlation would be equal to $+1$.

Graph for Problem 2.21



Problem 2.31



wife age	Husband age
20	22
25	27
28	30
30	32
35	37

↑
Made up data

Since $y = x + 2$ perfect ^{positive} linear relationship

therefore correlation should be
= 1