Homework #2 Solutions

1.59, 1.61, 2.11, 2.18, 2.20, 2.21

Problem 1.59

Salaries 25,000, 25,000, 25,000, 25,000, 25,000 (5 clerks)
60,000, 60,000 (2 junior accountants), 255,000

Mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$

$n = 8$

$\bar{X} = \frac{25,000 + 25,000 + 25,000 + 25,000 + 25,000 + 60,000 + 60,000 + 255,000}{8}$

$= 500,000/8 = 62,500$

All but the owner make less than the mean
ie 7 employers

Median is average of $\frac{8}{2} = 4$th and $\frac{8}{2}+1 = 5$th observations in sorted list ie 25,000

Problem 1.61

$$\bar{X} = \frac{25{,}000 + 25{,}000 + 25{,}000 + 25{,}000 + 25{,}000 + 60{,}000 + 60{,}000 + 455{,}000}{8}$$

$$= \$7{,}500$$ so the mean has increased

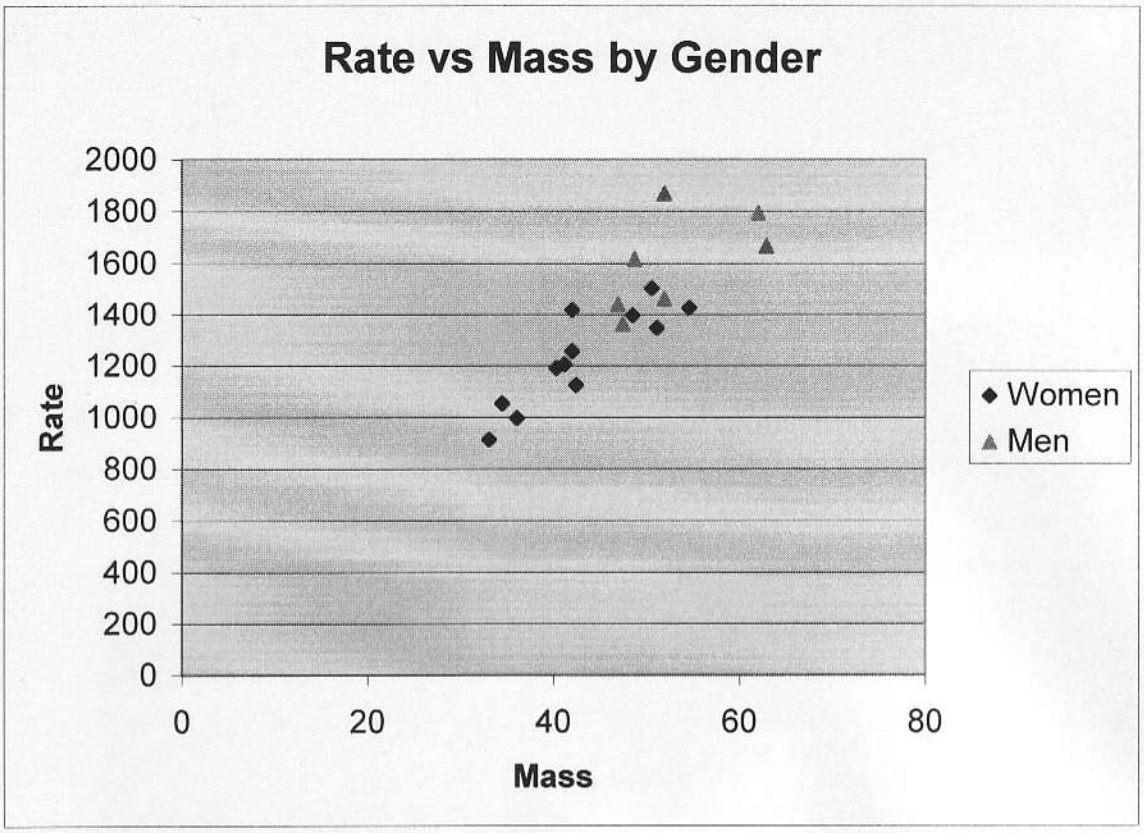
the median does not change since the 4th and 5th observations in the sorted list do not change

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

$$\text{mean Yellow} = \frac{45+51+48+46+38+47}{6}$$

$$= 47.16$$

$$\text{Mean white} = \frac{21+12+14+17+13+17}{6}$$

$$= 15.66$$

$$\text{Mean Green} = \frac{37+32+15+25+39+41}{6}$$

$$= 31.5$$

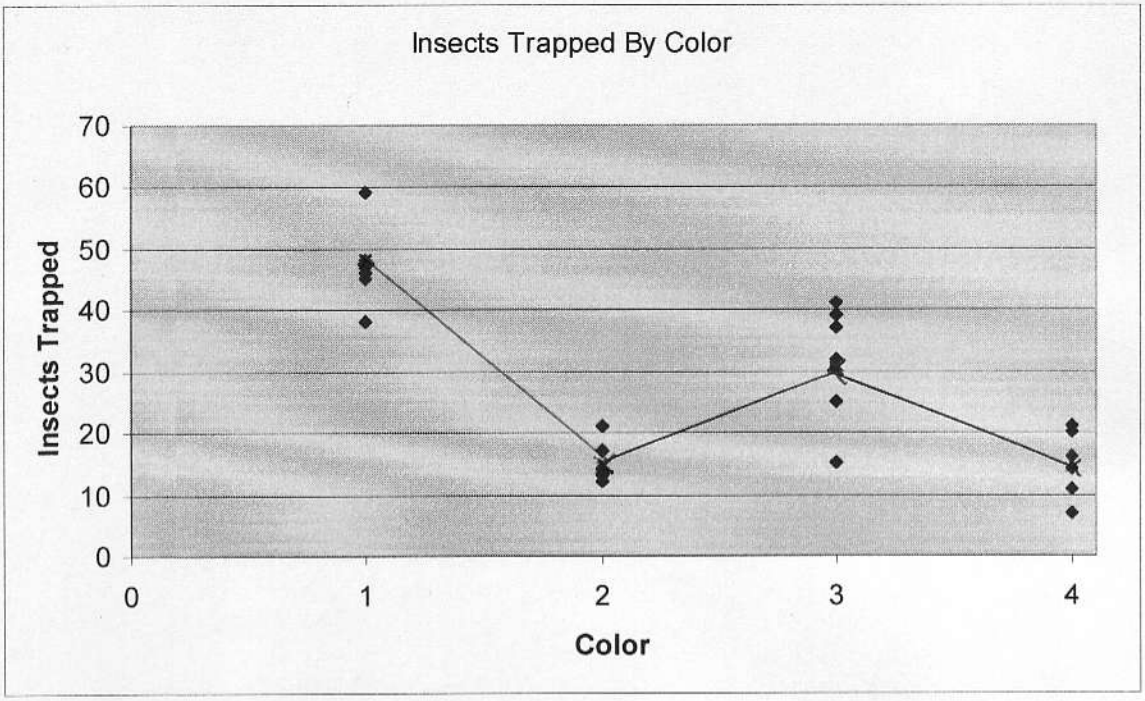
$$\text{mean blue} = \frac{16+11+20+21+14+7}{6}$$

$$= 14.83$$

(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 2.18



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 mean

Note X is Mass, y is rate

First for women

$n_{Women} = 12$

$$\sum_{i=1}^{12} x_i = 36.1 + 54.6 + 48.5 + 42.0 + 50.6 + 42.0 + 40.3 + 33.1 + 42.4 + 34.5 + 51.1 + 41.2$$

$$= 516.4$$

$$\sum_{i=1}^{12} y_i = 995 + 1425 + 1396 + 1419 + 1502 + 1256 + 1189 + 913 + 1124 + 1052 + 1347 + 1204$$

$$= 14821$$

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

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

$$\sum_{i=1}^{12} x_i^2 = 36.1^2 + 54.6^2 + \dots + 41.2^2$$
$$= 22741.34$$

$$\sum_{i=1}^{12} y_i^2 = 995^2 + 1425^2 + \dots + 1204^2$$
$$= 18695125$$

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

$$= 6.86$$

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

$$= 188.28$$

$$\sum_{i=1}^{12} x_i y_i = (36.1)(995) + (54.6)(1425) + \dots + (41.2)(1204)$$
$$= 650264.8$$

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

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

$$= 0.88 \ (2dp)$$

For men

$n = 7$

$$\sum_{i=1}^{7} x_i = 62 + 62.9 + \cdots + 46.9$$
$$= 371.7$$

$$\sum_{i=1}^{7} y_i = 1792 + \cdots + 1439$$
$$= 11200$$

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

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

$$\sum_{i=1}^{7} x_i^2 = 62^2 + 62.9^2 + \cdots + 46.9^2$$
$$= 20005.69$$

$$\sum_{i=1}^{7} y_i^2 = 1792^2 + 1666^2 + \cdots + 1439^2$$
$$= 18134870$$

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

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

$$\sum_{i=1}^{7} x_i y_i = (62.0)(1792) + (62.9)(1666) + \dots + (46.9)(1439)$$

$$= 599216.4$$

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

$$= \frac{1}{6}\left(\frac{1}{6.68}\right)\left(\frac{1}{189.23}\right)\left(599216.4 - 6(53.1)(1600)\right)$$

$$= 0.59 \text{ (2dp)}$$

(c) $\bar{x}_{women} = 43.03$

$\bar{x}_{men} = 53.1$

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

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

$= 66^2 + 64^2 + 66^2 + 65^2 + 72^2 + 65^2$

$= 26158$

$\sum_{i=1}^{6} y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 + y_6^2$

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

$= 28598$

5. $s_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{26158 - 6(66)^2}{6-1}}$

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

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

$s_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}} = \sqrt{\frac{28598 - 6(69)^2}{6-1}}$

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

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

6. $\sum_{i=1}^{6} x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 + x_5y_5 + x_6y_6$

$= (66)(72) + (64)(68) + (66)(70) + (65)(68) + (70)(71) + (65)(65)$

$= 27334$

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

from 1, from 5, from 6, from 3

$r = \frac{1}{6-1}\left(\frac{1}{2.0976}\right)\left(\frac{1}{2.5298}\right)\left(27334 - 6(66)(69)\right)$

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

$= 0.57 \ (2dp)$

(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