3a) $\frac{dQ}{dB_0} = 0 \rightarrow nb_0 + \Sigma_i X_i b_1 = \Sigma_i Y_i$

$Q = \sum_{i=1}^{n} (Y_i - B_0 - B_1 X_i)^2$

$\frac{dQ}{dB_0} = \frac{d \Sigma_i (Y_i - B_0 - B_1 X_i)^2}{dB_0}$

$\frac{dQ}{dB_0} = 0 = -2 \sum_i (Y_i - B_0 - B_1 X_i)$

values of B_0 & B_1 that mins Q is b_0 & b_1

so sub them in:

$0 = -2 \Sigma_i (Y_i - b_0 - b_1 X_i)$

$= \Sigma_i Y_i - nb_0 - \sum_{i=1}^{n} X_i b_1$

$nb_0 + \Sigma_i X_i b_1 = \Sigma_i Y_i$

3b) $\frac{dQ}{dB_1} = 0 \rightarrow \sum_i X_i b_0 + \sum_i X_i^2 b_1 = \sum_i X_i Y_i$

$$Q = \sum_{i=1} (Y_i - B_0 - B_1 X_i)^2$$

$$\frac{dQ}{dB_1} = 0 = -2 \sum_i X_i (Y_i - B_0 - B_1 X_i)$$

$$0 = \sum_i X_i Y_i - \sum X_i B_0 - \sum X_i^2 B_1$$

Sub b_0 & b_1 for B_0 & B_1

$$0 = \sum_i X_i Y_i - \sum_i X_i b_0 - \sum_i X_i^2 b_1$$

$$\sum_i X_i b_0 + \sum_i X_i^2 b_1 = \sum_i X_i Y_i$$

4a) show that $b_0 = \bar{Y} - b_1\bar{X}$

Given:

$\sum Y_i = nb_0 + \sum X_i b_1$ from problem 3a

$b_0 = \frac{1}{n}(\sum Y_i - \sum X_i b_1)$ | $\bar{Y} = \frac{1}{n}\sum Y_i$

$b_0 = \bar{Y} - b_1\bar{X}$ | $\bar{X} = \frac{1}{n}\sum X_i$

4b) show that $b_1 = \frac{SSXY}{SSX} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2}$

given from problem 3b $\qquad$ from 4a

$\sum X_i(Y_i - b_0 - b_1 X_i) = 0 \qquad b_0 = \bar{Y} - b_1\bar{X}$

$\sum X_i(Y_i - (\bar{Y} - b_1\bar{X}) - b_1 X_i) = 0$

$\sum X_i Y_i - \bar{Y}\sum X_i + b_1\bar{X}\sum X_i - b_1\sum X_i^2 = 0$

$\bar{Y} = \frac{1}{n}\sum Y_i$

$\bar{X} = \frac{1}{n}\sum X_i$

$b_1 \sum X_i^2 - \bar{X}^2 = \sum X_i Y_i - \frac{\sum X_i Y_i}{n}$

$\sum X_i Y_i - \frac{\sum X_i Y_i}{n} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$

$b_1 = \frac{\sum X_i Y_i - \frac{\sum X_i Y_i}{n}}{\sum (X_i - \bar{X})^2}$

$SSXY = \sum (X_i - \bar{X})(Y_i - \bar{Y})$

$SSE = \sum (X - \bar{X})^2$

$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{SSXY}{SSE}$

1.12a)

Observational

1.12 b)

While increasing the time spent exercising could help decrease the amount of colds, that conclusion is too simple & there are other variables that play a role

1.12 c)

1) Age of senior citizens
2) Pre-existing medical conditions
3) diet
4) medication

1.12d)

1) Remove any Senior Citizen's that have a pre-existing conditions (as they would be outliers)

2) Separate the senior citizens by age group (as younger citizens should be healthier & have less colds than older)

3) Have each participant eat a relatively healthy diet

1.13a)

Observational

1.13b)

While increasing the time in class prep could help productivity levels increase, that conclusion is too simple & there are other variables that play a role

1.13c)

1) Number of homework/extra problems completed
2) Number of hours spent outside of class to understand topics

1.13d)

1) Record number of problems/hw completed

2) Record number of hours spent outside of class learning about topics covered in class

3) Group employees by these extra variables & find statistical relationship

1.16)

False, the residuals must have a normal distribution, not Y

1.17)

correct, method of least-squares can be used to estimate b_0 \& b_1. We see this using:

$$Q = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

1.19a)

$$\bar{X} = 24.725 \cong 24.7$$

$$\bar{Y} = 3.07405 \cong 3.1$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 92.40565 \cong 92.4$$

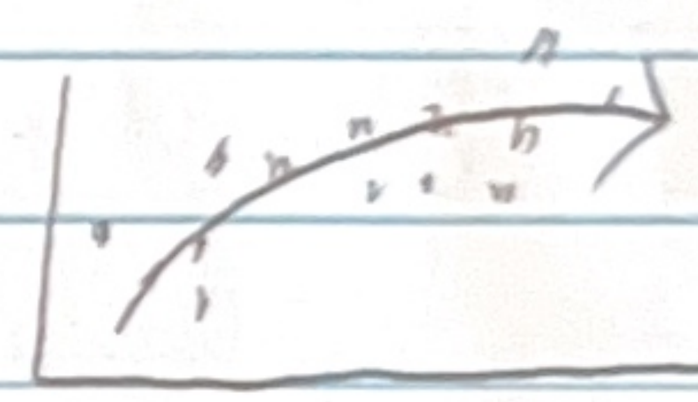
$$\sum (X_i - \bar{X})^2 = 2,379.925 \cong 2,379.3$$

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})} = \frac{92.4}{2,379.3} = 0.039$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 3.1 - (0.039)(24.7) = 2.14$$

1.19a) $\hat{Y} = b_0 + b_1 X \rightarrow \hat{Y} = 2.14 + 0.039X$

1.19b)
It looks like the estimated regression fits it fine, but a better estimate would seem to be one with a square root slope as the points seem to curve:



1.19c) $\hat{Y} = 2.14 + 0.039(30)$

$= 3.31$

1.19d)

$2.14 + 0.039(31) = 3.349$

$3.349 - 3.31 = 0.039$

1.21(a) $b_0 = 10.2$

$b_1 = 4$

$\hat{Y} = 10.2 + 4X$

1.21(b)

$10.2 + 4(1) = 14.2$

1.21(c)

$10.2 + 4(2) = 18.2$

$18.2 - 14.2 = 4$

1.21(d)

$\bar{X} = 10/10 = 1$

$\bar{Y} = 142/10 = 14.2$

$(\bar{X}, \bar{Y}) = (1, 14.2)$

Y= length of stay
X_1= infection risk
X_2= available facilities & services
X_3= routine chest X-ray ratio

1.45 a)

$$\hat{Y}_1 = 6.34 + 0.76X_1$$

$$\hat{Y}_2 = 7.72 + 0.04X_2$$

$$\hat{Y}_3 = 6.57 + 0.04X_3$$

1.45 b)

A linear relation seems to fit the infection risk well

A linear relation seems to fit available facilities & services well

A linear relation seems to fit routine chest X-ray ratio well

1.45c) for infection Risk

$n = 113$

$$\hat{\sigma}^2_{x_1} = \frac{\sum(Y_i - \hat{Y}_i)^2}{n}$$

$$= \frac{-0.156^2}{113} = 2.15 \times 10^{-4}$$

$$MSE_{x_1} = \frac{113}{113-2}(2.15 \times 10^{-4}) = 2.19 \times 10^{-4}$$

---

for available facilities

$$\hat{\sigma}^2_{x_2} = \frac{22.82^2}{113} = 4.61$$

$$MSE_{x_2} = \frac{113}{111}(4.61) = 4.69$$

---

for routine chest X-ray ratio

$$\hat{\sigma}^2_{x_3} = \frac{-21.11^2}{113} = 3.94$$

$$MSE_{x_3} = \frac{113}{111}(3.94) = 4.01$$

Infection Risk leads to the smallest variability around the fitted regression line