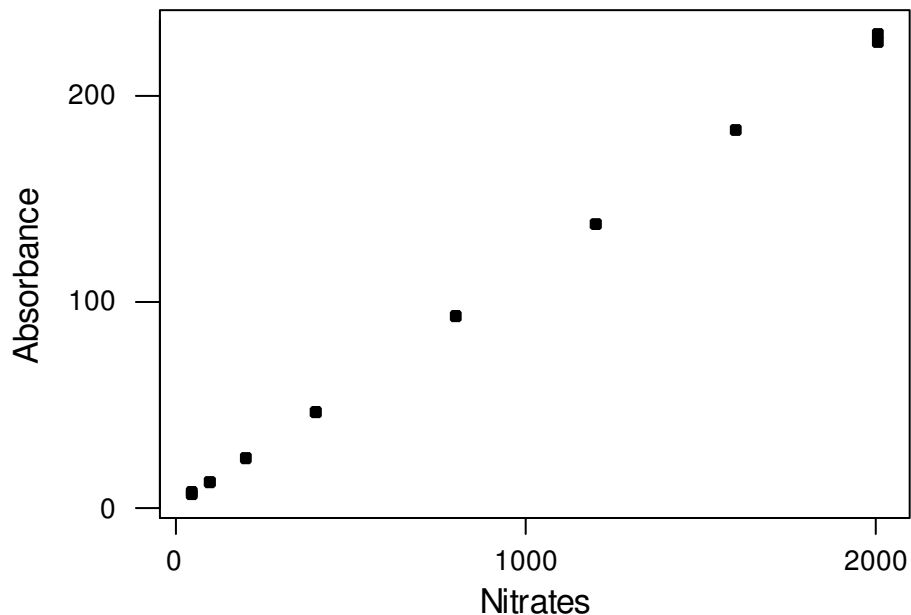


## Statistics 528 – Homework 4 Solutions

2.40

(a)



Minitab gives the correlation 1.0 (in fact, .99994). Note that it is rounded up.  
So, recalibration would not be necessary.

(b) The equation of the least squares line is  $\text{Absorbance} = 1.66 + 0.113 \text{ Nitrates}$ .  
At 500mg of nitrates per liter, the predicted absorbance is  $58.16 (= 1.66 + 0.113 * 500)$ . Because the correlation is very close to 1, we expect that the prediction would be very accurate.

2.46

(a) Plot below.

(b) Correlation is 0.503. R-square is 0.253. The correlation 0.503 suggests that there is a positive linear association between U.S. returns and overseas returns. However, the strength of the linear relation is not high enough to predict overseas returns accurately based on U.S. returns. The R-square 0.253 indicates that only 25.3% of variation in overseas returns can be explained by linear regression on U.S. returns.

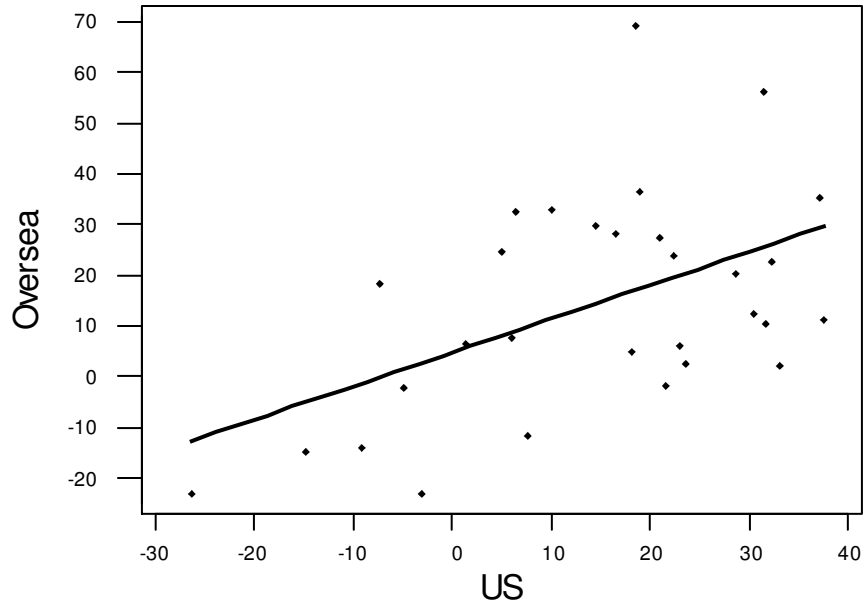
(c)  $\text{Overseas} = 4.76 + 0.663 \text{ US}$ . For 1993, the predicted value is  $11.46 (= 4.76 + 0.663 * 10.1)$ , and the actual value is 32.9.

(d) Predictions will not be very accurate because the correlation is low.

## Regression Plot

$$\text{Oversea} = 4.75834 + 0.662769 \text{ US}$$

S = 19.0417    R-Sq = 25.3 %    R-Sq(adj) = 22.7 %



2.52

- (a) The slope is  $b=0.6 \cdot 8/30=0.16$ . The intercept is  $a=75-0.16 \cdot 280=30.2$ .
- (b) Her predicted score is  $30.2+0.16 \cdot 300=78.2$ .
- (c)  $R\text{-square}=0.6 \cdot 0.6=0.36$ , which is fairly low. This means that scores can be quite a bit higher or lower than the regression line.

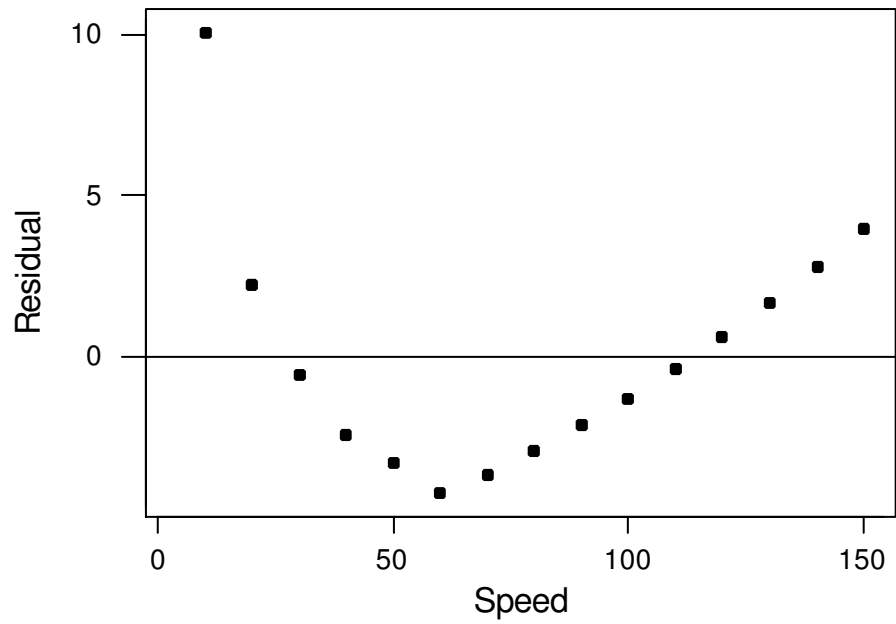
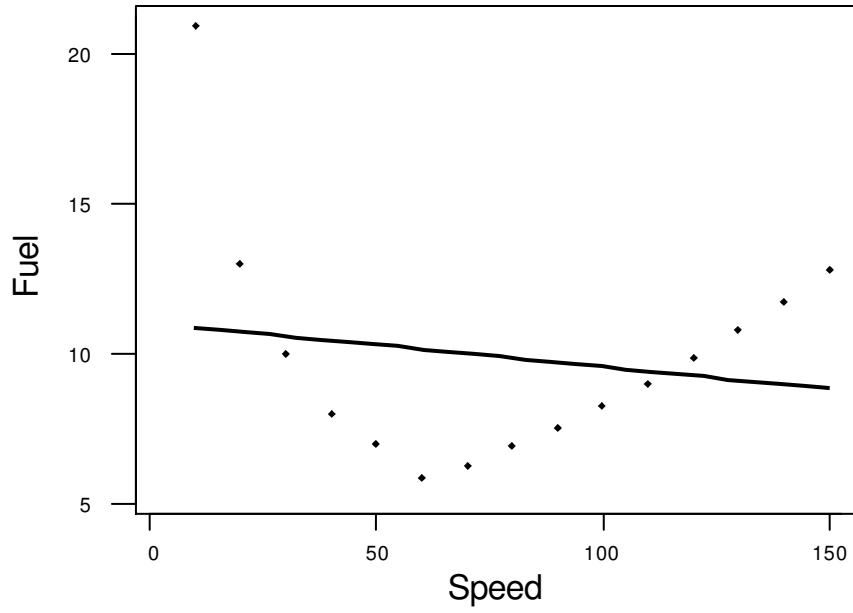
2.58

- (a) Plot below;
- (b) No. There is an obvious pattern in the data that is not linear. The pattern cannot be described by a straight line, thus using the regression line to predict  $y$  from  $x$  would not be appropriate.
- (c) The residuals sum to  $-0.01$ . (Not 0, because of rounding error.)
- (d) Residual plot is second below.

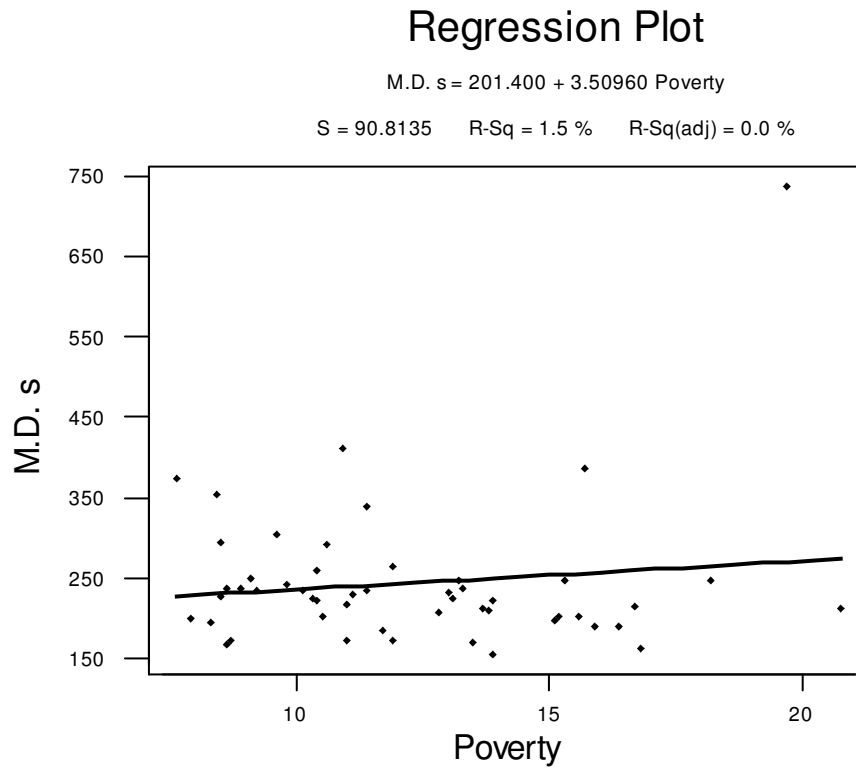
# Regression Plot

$$\text{Fuel} = 11.0579 - 0.0146571 \text{ Speed}$$

S = 3.90475    R-Sq = 2.9 %    R-Sq(adj) = 0.0 %



2.64  
(a)



The regression line using all 51 values is the one with (slightly) positive slope.

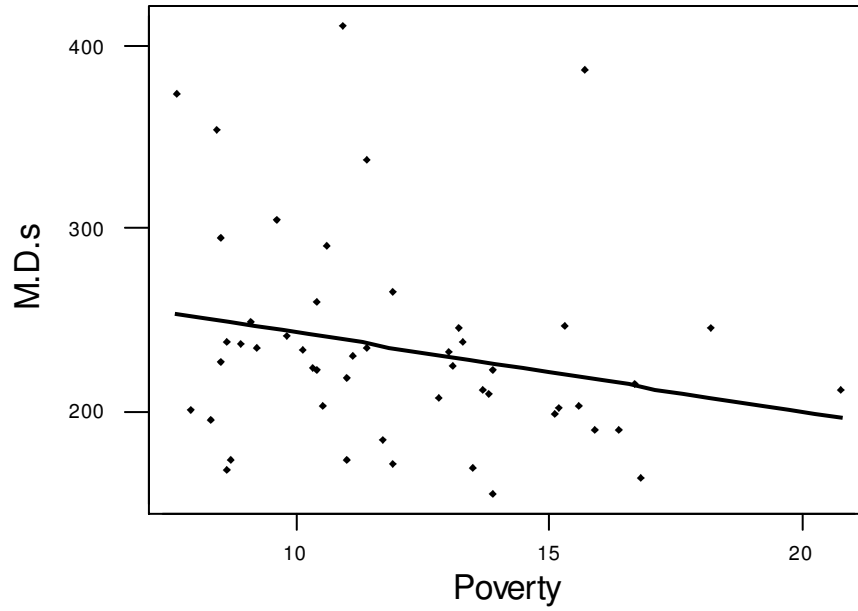
(b) DC is the second largest horizontally and the largest vertically. A point this extreme in X, as well as extreme in Y will be influential.

(c) Omitting DC produces the regression line below that has negative slope, so DC was, indeed, influential.

# Regression Plot

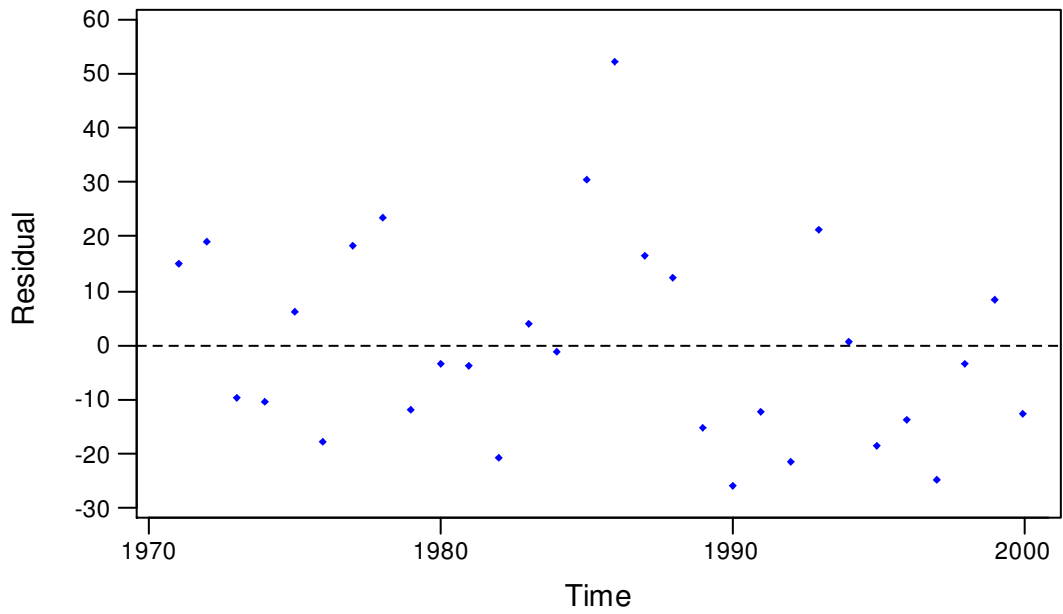
M.D.s = 287.035 - 4.37524 Poverty

S = 56.6442    R-Sq = 5.3 %    R-Sq(adj) = 3.4 %



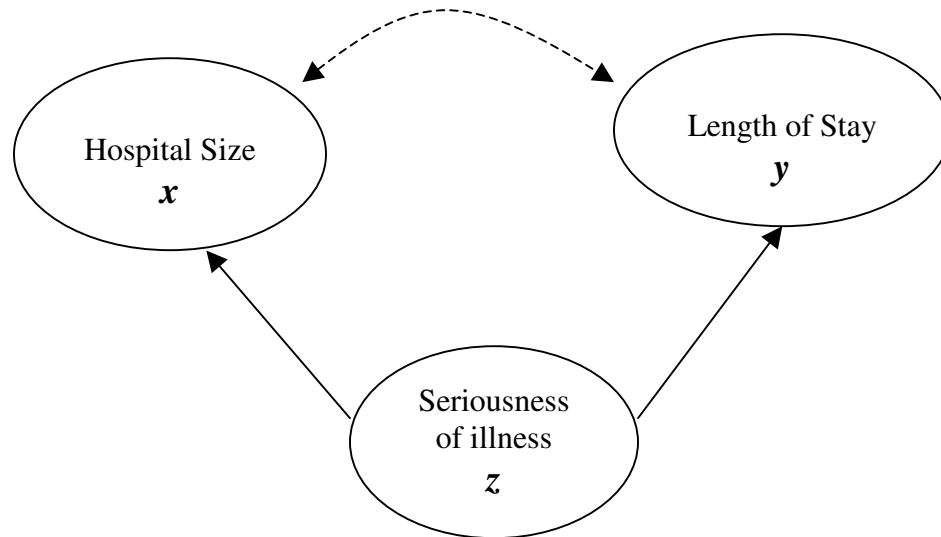
# Residuals Versus Time

(response is Oversea)



2.84

This spurious positive correlation between  $x$  and  $y$  can be explained by the presence of a lurking variable such as seriousness of illness which influences both  $x$  and  $y$ .



2.86

A better explanation is that people who are heavier use artificial sweeteners in order to lose weight.