$\boxed{\text{3.2 HW (DAY 2)}}$ #'s 35, 37, 39, 41, 43, 45

**35** New bar - 80 grams decreases 6 grams / day

OPTION 1

REGRESSION EQ and must define $\hat{y}$ + x

$\hat{y} = 80 - 6x$ where
$\hat{y}$ = estimated soap weight
x = # days since bar was new

OPTION 2

Use words in EQUATION

$\widehat{\text{Soap Weight}} = 80 - 6(\text{DAYS})$

**37**

$\widehat{\text{highway mpg}} = 4.62 + 1.109(\text{city mpg})$

4.62 ↑ y int, 1.109 ↑ slope

a) The slope is 1.109. We predict that highway mileage will increase by 1.109 mpg for each 1 mpg increase in city mileage

b) The y intercept is 4.62 mpg. This is NOT statistically meaningful because this would represent the highway mileage for a car that gets 0 mpg in the city

c) Car = 16 city mpg

$\widehat{\text{highway}} = 4.62 + 1.109(16) = 22.36$ / The predicted highway mileage is 22.36 mpg

Car = 28 city mpg

$\widehat{\text{highway}} = 4.62 + 1.109(28) = 35.67$ The predicted highway mileage is 35.67 mpg

3.2 cont

39 $\hat{pH} = 5.43 - 0.0053(\text{weeks})$

(a) The slope is $-.0053$; the pH decreases by .0053 units per week ON AVERAGE.

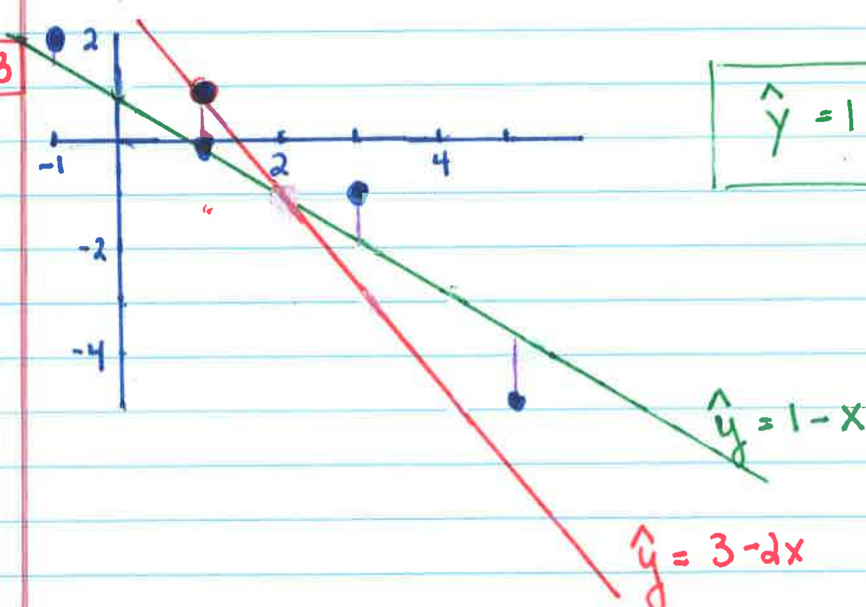
(b) The y intercept is 5.43 and it estimates the Ph level at the start of the study.

(c) At the end of the study, the predicted Ph was 4.635

$$\hat{pH} = 5.43 - .0053(150) = 4.635.$$

41 It would be inappropriate to predict the PH after 1,000 months. One thousand months corresponds to about 4,000 weeks, which is well outside the observed time period of 150 weeks. This constitutes EXTRAPOLATION.

43




$\hat{y} = 1 - x$ fits the data (5 points) better because the line is closer to the points

3.2 cont

45 Actual pH = 5.08 at week 50

$\hat{pH} = 5.43 - .0053(50)$  $\hat{pH} = 5.165$

residual = $y - \hat{y} = 5.08 - 5.165 = -.085$

The residual is $-0.085$ meaning that the line predicted a pH value for that week that was .085 too large.

# 3.2 DAY 3 #47, 49, 54, 56, 58

47 Men's Height $\bar{y} = 68.5$ in $S_y = 2.7$ in
Women's Height $\bar{x} = 64.5$ in $S_x = 2.5$ in

$\boxed{r = .5}$

See your green sheet for formula's

$\hat{y} = b_0 + b_1 x$

$b_0 = \bar{y} - b_1 \bar{x}$

$b_1 = r \frac{S_y}{S_x}$

(a) $\widehat{\text{men}} = \cancel{a}^{b_0} + \cancel{b}^{b_1}(\text{women})$

$b_1 = .5\left(\frac{2.7}{2.5}\right)$

$\boxed{b_1 = .54}$

$b_0 = 68.5 - .54(64.5)$

$\boxed{b_0 = 33.67}$

LSRL: Option 1: $\widehat{\text{men}} = 33.67 + .54(\text{women})$

OR

Option 2: $\hat{y} = 33.67 + .54x$

where $x$ = women's height
$y$ = men's height

(b) Women's height = 67 in (1 SD ABOVE MEAN)

Think about this EQ: $\boxed{b = r \frac{S_y}{S_x}} = \frac{(.5)(2.7)}{(2.5)}$
(for slope)

FOR EACH ADDITIONAL 2.5 in in women's height we expect men's height to change by $r \cdot S_y$ inches

Therefore since women's height increased by 1 SD men's height is

$\bar{y} + rS_y = 68.5 + .5(2.7)$

predicted men's height is 69.85 in

PG 173

3-2 Cont

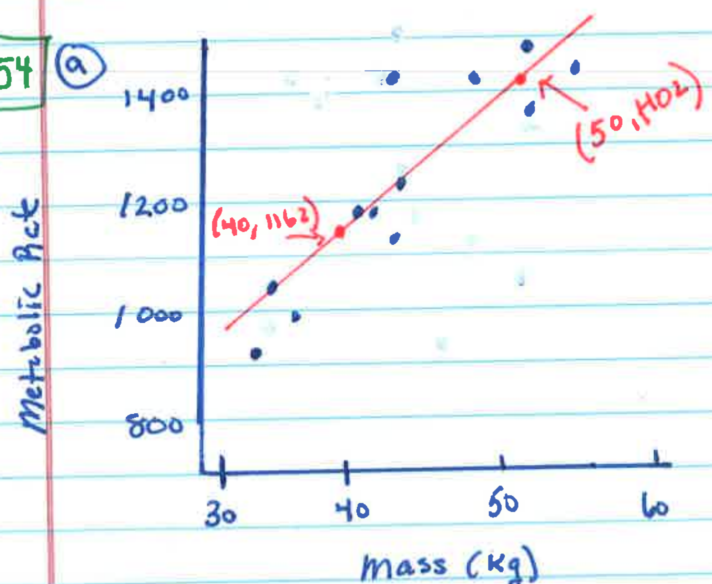
49 (a) $r^2 = (.5)^2 = .25$ $\boxed{r^2 = .25}$

$r^2 = .25$ which means the straight line relationship, explains 25% of the variation in husband's height

Pg 177

(b) s = 1.2in is the standard deviation of the RESIDUALS. This value is the typical or average prediction error when using this line for prediction.

54 (a)




(b) $\hat{y} = 201.2 + 24.026X$

x = mass (kg)

y = metabolic rate

Tip ENTER EQ

y =

Vars

> 5: statistics

> EQ

> 1: REGEQ

GO TO Table

pick 2 points to plot line

OR USE TRACE

(c) The slope tells us that we would predict an increase in the metabolic rate of about 24 cal/day for each additional kilogram of body mass

(d) X = 45 Kg $\widehat{Rate} = 201.2 + 24.026(45)$

$\widehat{Rate} = 1282.4$

The predicted metabolic rate is 1,282.4 cal/day

(3.2 CONT)

56 (a) To create a residual plot - make a Scatter plot STAT PLOT

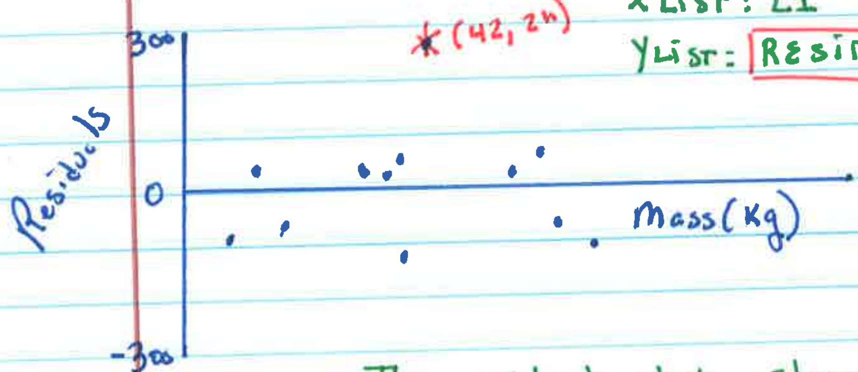
X LIST : L1

Y LIST : RESID ← 2ND LIST

> NAMES

> 7: RESID

ZOOM > 9




The residual plot shows NO PATTERN AND there fore the linear fit is good. There is 1 large positive residual. The outlier is near the mean of mass value ($\bar{x} \approx 43$ Kg), it does NOT influence the line very much

(b) The point with the largest residual of about 200 means that the line greatly under predicted the metabolic rate for this person

58 $r^2 = .768$ 76.8% of the variation in the metabolic rate is explained by the straight line relationship.

$S = 95.08$ The average error (residual) when using the line for prediction is 95.08 Calories burned per 24 hours.

3.2 HW DAY 4

#'s 59, 60, 61, 63, 65, 67, 71 – 78

59 (a) There is a strong, positive, linear association between the lab measurements and field measurements. There is more variation in field measurements for larger lab measurements

(b) The points for the larger depts fell systematically below the line $y = x$ showing that the field measurements are too small compared to the lab measurements.

(c) To fit the data, the LSRL would be pulled down to fit the larger lab measurements which would result in the slope decreasing. And the y intercept would increase.


60 The residual plot clearly shows that the prediction error increases for the larger lab measurements

3.2 cont

61 The residual plot shows a clear curved pattern. Therefore this LSRL would NOT be an appropriate model for this data.

63 (a) MINITAB

| Predictor | Coef | |
|---|---|---|
| Constant | 157.68 | ← y intercept |
| Pairs | -2.9935 | ← slope |

$\hat{y} = 157.68 - 2.99x$ where $x$ = # of breeding pairs, $y$ = % males returning

OR $\widehat{\text{RETURNING}} = 157.68 - 2.99(\text{PAIRS})$

$X = 30 \longrightarrow \widehat{\text{Returning}} = 158.68 - 2.99(30) = 68.98$

We predict that about 69% of males will return, for a season with 30 breeding pairs.

(b) $\boxed{r^2 = 63.1\%}$ The linear relationship explains 63.1% of the variation in the percent of returning males

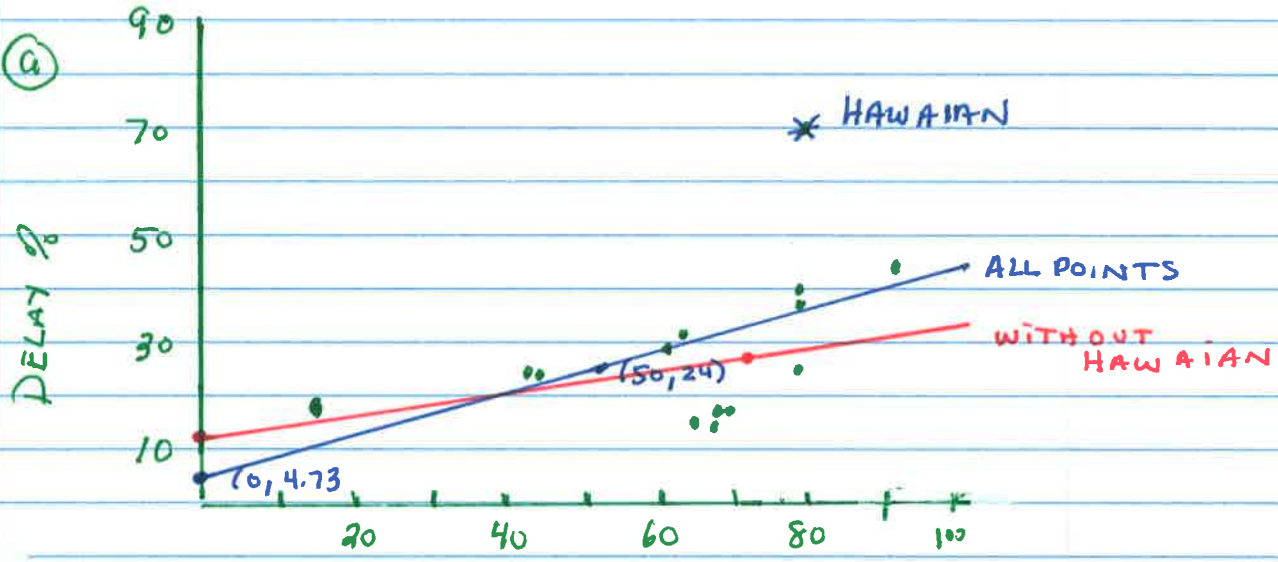
(c) $r = \sqrt{.631} = .794$ $\boxed{r = -.79}$ The sign is negative because it is the same sign as the slope.

(d) $\boxed{S = 9.46334}$

The typical error when using this line to predict the return rate of males is about 9.46%.

3-2 Cont

65 (a)




ALL POINTS

$r = .4765$   $\boxed{\widehat{delay} = 4.73 + .3868(\text{OUTSource})}$

remove HAWAIAN (80, 70)   $r = .4838$   $\boxed{\widehat{delay} = 10.88 + .2495\,(\text{OUT Source})}$

(b) The correlation for all points is $r = .4765$. It rises slightly to .4838 removing Hawaian. This outlier has too small a change to consider the outlier influential based on correlation.

(c) FOR AN AIRLINE WITH $X = 76$ (outsourced 76%) PREDICTIONS USING THE 2 LSRL's are:

all points   $\widehat{delay} = 4.73 + .3868(76) = 34.13$   prediction 34.13% delay

removing Hawaian   $\widehat{delay} = 10.88 + .2495(76) = 29.84$   prediction 29.84% delay

This shows a substantial difference in predictions indicating that the outlier is influential for regression. In addition the LSRL slopes and y intercepts are also very different supporting the outlier is influential for regression.

3.2 Cont    MC #'s 71 – 78

[71] (B) Look at graph for $x = 110 \rightarrow y \sim 60$

[72] (C) The slope is positive so this limits to 1, 2, 4, 6

pts (90, 45) (110, 60)

$m = \Delta y / \Delta x = \frac{45-60}{90-110} = \frac{-15}{-20} = .75 \cong 1$

[73] (b) There is a negative association between smoking and life expectancy

[74] (a) slope = .93

[75] (b) $\hat{y} = 6.4 + .93(100) = \underline{\underline{99.4}}$

[76] (a) Correlation and slope have the same sign since slope = .93 then the corr is positive

[77] (d) $r^2 = .95$ measures variation accounted by LSRL

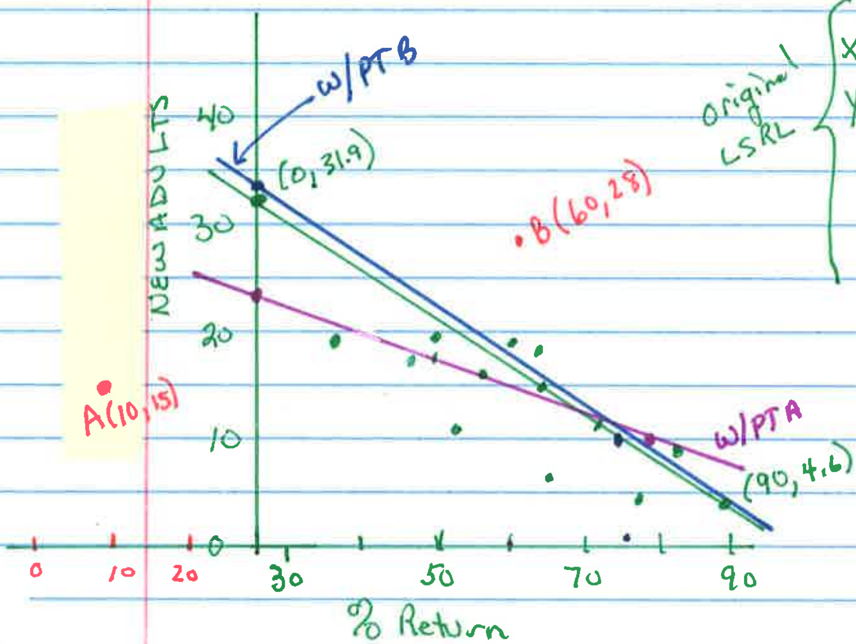
[78] (a) $\hat{y} = 6.4 + .93(60) = 62.2$    x = armspan = 60

residual = $y - \hat{y} = 59 - 62.2 = -3.2$    y = height = 59

3-2 Cont

$\boxed{67}$ ORIGINAL DATA — ALWAYS CREATE SCATTER PLOT

FIND means & STD DEV.




original LSRL

X (% Return) $\bar{x} = 58.2$ $s_x = 13.0$

Y (Adults) $\bar{y} = 14.2$ $s_y = 5.3$

$\widehat{\text{Adults}} = 31.93 - .304(\%\text{ Return})$

$r = -.75$ $r^2 = .56$

(a) $\boxed{\text{PT A}}$ is A HORIZONTAL OUTLIER

$\boxed{\text{PT B}}$ is a Vertical outlier

(b) original plus pt A $\widehat{\text{Adults}} = 22.8 - .156(\%\text{ Return})$

$r = -.55$ $r^2 = .30$

original plus pt B $\widehat{\text{Adults}} = 32.3 - .293(\%\text{ Return})$

$r = -.58$ $r^2 = .34$

CONCLUSION

1. Adding point B has little impact on the regression line. The slopes are similar (-.304 and -.293) and the y intercepts are similar (31.93 and 32.3). The line has shifted up slightly to pull towards point B. Also note that pt B's X coordinate is close to $\bar{X}$.
2. Point A is an influential point and dramatically pulls the line down towards the point. Both the slope and y intercept are very different TO THE ORIGINAL EQUATION