Chapter 10-Regression

10.1 Regression equation predicting infant mortality from income

Y = Infant mortality
X = Income
$\overline{Y}$ = 6.70 s_Y = 0.698 s_Y^2 = 0.487
$\overline{X}$ = 46.00 s_X = 6.289 s_X^2 = 39.553
cov_{XY} = 2.7245

$$b = \frac{\text{cov}_{XY}}{s_X^2} = \frac{2.7245}{39.553} = 0.069$$

$$a = \overline{Y} - b\overline{X} = 6.70 - (0.069)(46.00) = 3.53$$

$$\hat{Y} = 0.069(X) + 3.53$$

10.3 If the high risk fertility rate jumped to 70, we would predict that the incidence of birthweight < 2500gr would go to 8.35.

$$\hat{Y} = bX + a = 0.0689X + 3.53$$

$$= 0.0689 * 70 + 3.53 = 8.35$$

This assumes that there is a causal relationship, which is plausible in some ways, but not proven.

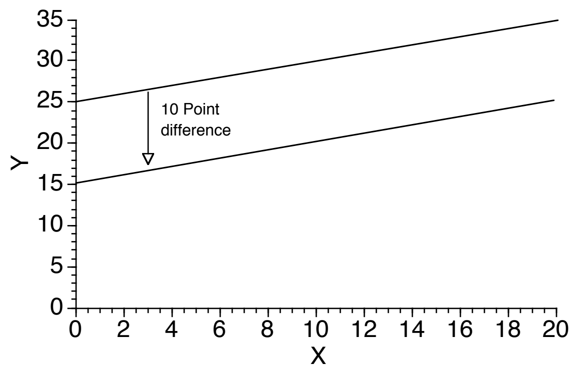
10.5 I would be more comfortable speaking about the effects on Senegal because it is already at approximately the mean income level and we are not extrapolating for an extreme country.

10.7 Prediction of Symptoms score for a Stress score of 45:

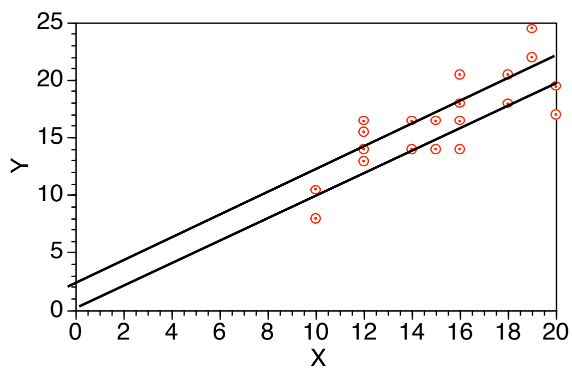
Regression equation: $\hat{Y}$ = 0.7831X + 73.891
If X = 45: = 0.7831*45 + 73.891
Predicted Symptoms = 109.13

10.9 Subtracting 10 points from every X or Y score would not change the correlation in the slightest. The relationship between X and Y would remain the same.

10.11 Diagram to illustrate Exercise 10.10:



10.13 Adding a constant to *Y*:



a) From this figure you can see that adding 2.5 to *Y* simply raised the regression line by 2.5 units.
b) The correlation would be unaffected.

10.15 Predicting GPA (*Y*) from ADDSC (*X*):

$$b = \frac{\text{cov}_{XY}}{s_X^2} = \frac{-6.580}{154.431} = -0.0426$$

$$a = \overline{Y} - b\overline{X} = 2.456 - 0.0426 * 52.602 = 4.699$$

$$\hat{Y} = -0.0426X + 4.699$$

10.17 For the faculty, the starting salary would appear to be (on average) \$31,000. This is the intercept, and therefore the value of salary when years of service is 0. Salary seems to increase at 900 for every year of service (which, for the new faculty member is a 3% raise, but it is considerably smaller than that for more senior faculty). Administrators average a much lower starting salary (\$18,000), but they increase \$1,500 for every year of service.

One way to solve for the point at which they become equal is to plot a few predicted values and draw regression lines. Where the lines cross is the point at which they are equal. A more exact way of to set the two equations equal to each other and solve for *X*.

$$0.9X + 31 = 1.5X + 18$$

Collecting terms we get

$$31 - 18 = 1.5X - 0.9X$$

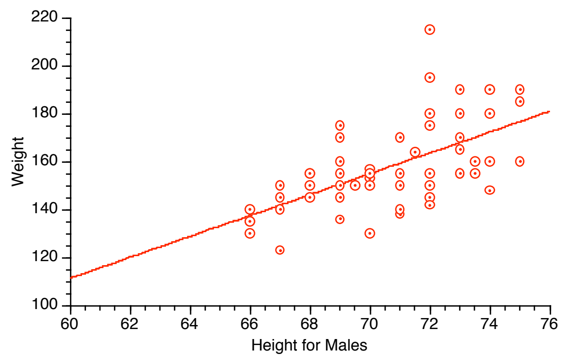
$$13 = 0.6X$$

$$X = 13/0.6 = 21.67$$

To check this, substitute 21.67 in both equations

$$0.9 * 21.67 + 31 = 50.503 = 1.5 * 21.67 + 18$$

10.19 Weight as a function of height for males:



The regression solution that follows is a modification of printout from SPSS.

```
Equation Number 1     Dependent Variable..   WEIGHT
Variable(s) Entered on Step Number
   1..     HEIGHT
Multiple R            .60368
R Square              .36443
Adjusted R Square     .35287
Standard Error      14.99167

Analysis of Variance
                    DF       Sum of Squares      Mean Square
Regression           1           7087.79984       7087.79984
Residual            55          12361.25279        224.75005

F =      31.53637       Signif F =  .0000

------------------ Variables in the Equation ------------------
Variable                B          SE B        Beta          T  Sig T

HEIGHT          4.355868      .775656    .603680      5.616  .0000
(Constant)   -149.933617    54.916943                -2.730  .0085
```

b) The intercept is given as the "constant" and is -149.93, which has no interpretable meaning with these data. The slope of 4.356 tells us that a one-unit increase in height is associated with a 4.356 increase in weight.
c) The correlation is .60, telling us that for females 36% of the variability in weight is associated with variability in height.
d) Both the correlation and the slope are significantly different from 0, as shown by an *F* of 31.54 and a (equivalent) *t* of 5.616.

10.21 Predicting my own weight, for which I use the equation from Exercise 10.19:

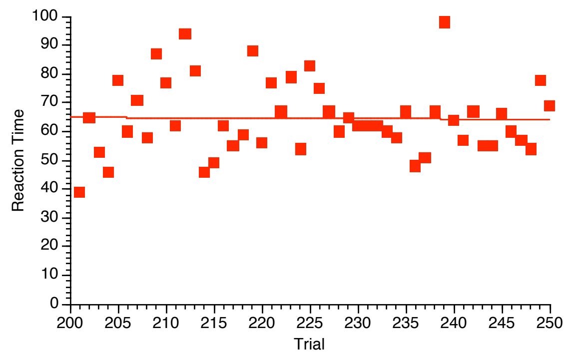
$\hat{Y}$ = 4.356*height - 149.93
$\hat{Y}$ = 4.356*68 - 149.93 = 146.28

a) The residual is Y - $\hat{Y}$ = 156 - 146.28 = 9.72. (I have gained some weight since I last used this example.)
b) If the students who supplied the data gave biased responses, then, to the degree that the data are biased, the coefficients are biased and the prediction will not apply accurately to me.

10.23 Predictions for a 5'6" male and female

For the male, $\hat{Y}$ = 4.356*66 - 149.93 = 137.57
For a female, $\hat{Y}$ = 2.578*66 - 44.859 = 125.29
Difference = 12.28 pounds

10.25 Plot of Reaction Time against Trials for only the Yes/5-stimuli trials:



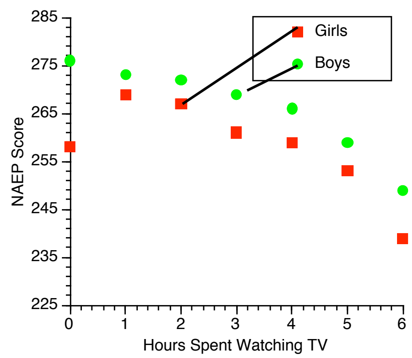
The following regression solution is a modification of SPSS printout.

```
Equation Number 1    Dependent Variable..   RXTIME
Variable(s) Entered on Step Number
   1..     TRIAL

Multiple R           .01640
R Square             .00027
Adjusted R Square   -.02056
Standard Error     12.76543
```

```
Analysis of Variance
                    DF       Sum of Squares       Mean Square
Regression           1              2.10363           2.10363
Residual            48           7821.89637         162.95617

F =        .01291       Signif F =  .9100

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta          T  Sig T

TRIAL          -.014214      .125100   -.016397      -.114  .9100
(Constant)    67.805186    28.267795                 2.399  .0204
```

The slope is only -0.014, and it is not remotely significant. For this set of data we can conclude that there is not a linear trend for reaction times to change over time. From the scatterplot above we can see no hint that there is any nonlinear pattern, either.

10.27 The evils of television:



Regression equations:

Boys $\hat{Y} = -4.821X + 283.61$

Girls $\hat{Y} = -3.460X + 268.39$

b) The slopes are roughly equal, given the few data points we have, with a slightly greater decrease with increased time for boys. The difference in intercepts reflects the fact that the line for the girls is about 9 points below that for boys.

c) Television can not be used as an explanation for poorer scores in girls, because we see that girls score below boys even when we control for television viewing.

10.29 Draw a scattering of 10 data points and drop your pencil on it.

b) As you move the pencil vertically you are changing the intercept.

c) As you rotate the pencil you are changing the slope.

d) You can come up with a very good line simply by rotating and raising or lowering your pencil so as to make the deviations from the lines as small as possible. (We really minimize squared deviations, but I don't expect anyone's eyes to be good enough to do that.)