Simple Linear Regression

11.1 Creating the Least Squares Equation

1. Researchers looked at the link between skipping breakfast and obesity. For each of eleven randomly selected female patients researchers recorded the number of days per week that breakfast is typically eaten and their BMI. Use the data to find the least square prediction line with BMI as the response variable.
(Note: $\sum X = 34, \sum Y = 304, \sum XY = 856, \sum X^2 = 178$)

Subject	1	2	3	4	5	6	7	8	9	10	11
Days w/Breakfast	0	2	1	0	7	7	3	5	6	1	2
BMI	30	27	31	35	24	23	27	25	25	28	29

2. Medical researchers studied the connection between waist circumference and HDL (good) cholesterol in children. The data below if from a sample of ten randomly selected children. Use the data to create the least squares prediction line with HDL cholesterol as the response variable. Would it be a good idea to use the model to estimate the average HDL level for children with waist circumference equal to 105 cm?
(Note: $\sum X = 737.4, \sum Y = 499, \sum XY = 36,469.53, \sum X^2 = 55,106.24$)

Subject	1	2	3	4	5	6	7	8	9	10
WC	64.6	69.7	91.3	63.0	68.1	74.1	73.8	86.2	70.2	76.4
HDL	56.1	53.0	44.3	55.2	53.4	47.6	46.8	44.9	52.1	45.6

3. A college president wants to know if a link exists between reading classic literature in high school and having a high grade point average in college. He has the statistics department collect the following data. Use the data collected to find the least squares regression line that uses number of classics read in high school as a predictor for GPA. What is the predicted average GPA for students who read 10 classics in high school? Do any of the ordered pairs of data in the data set stand out as being very different from the others? If so, what effect might this have on our model?
(Note: $\sum X = 102, \sum Y = 28.1, \sum XY = 341.2, \sum X^2 = 1526$)

Student	1	2	3	4	5	6	7	8	9
NOCR	16	6	7	5	8	14	4	20	22
GPA	3.3	2.5	2.6	2.2	3.0	3.1	3.7	3.7	4.0

Answers:

1. $\hat{y} = 31.18 - 1.15x$

2. $\hat{y} = 82.89 - 0.45x$

   It's a bad idea to use the model for a waist circumference of 105 cm since the data used to create the model only ranged from 63.0 to 91.3 cm. A linear model may not be appropriate for the values outside of this range.

3. $\hat{y} = 2.43 + 0.06x$

   Plugging ten into our models gives us an average GPA of 3.03.

   ($\hat{y} = 2.43 + 0.06(10) = 3.03$)

   Student number 7 has an unusual value compared to the others in the list. This value may be called an outlier. It does not follow the pattern in the rest of the data. It will have the effect of increasing the SSE of the model. You can see it stands out in the graph below .