Simple Linear Regression

11.2 Finding S for the Random Error Terms

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 S^2 the sample variance for the random error terms of the model.
(Note: $\sum X = 34, \sum Y = 304, \sum XY = 856, \sum X^2 = 178, \sum Y^2 = 8524$)

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 find S^2 the sample variance for the random error terms of the model. (Note:
$\sum X = 737.4, \sum Y = 499, \sum XY = 36,469.53, \sum X^2 = 55,106.24, \sum Y^2 = 25,083.08$)

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 below. Use the data to find S^2 the sample variance for the random error terms of the model.
(Note: $\sum X = 102, \sum Y = 28.1, \sum XY = 341.2, \sum X^2 = 1526, \sum Y^2 = 90.73$)

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

4. In problem 2 above, we found S^2 = 4.602 for the data used to predict the average HDL cholesterol level for children given their waist circumference. Find S and then state the largest deviation we would expect between any of the actual data points and our least squares line.

Answers:

1. The preliminary calculations and final answer:

$$SS_{xx} = 72.9091, SS_{xy} = -83.6364, SS_{yy} = 122.5455$$
$$\hat{\beta}_1 = -1.1471$$
$$SSE = 26.6035$$
$$S^2 = 2.956$$

2. The preliminary calculations and final answer:

$$SS_{xx} = 730.364, SS_{xy} = -326.73, SS_{yy} = 182.98$$
$$\hat{\beta}_1 = -0.4474$$
$$SSE = 36.817$$
$$S^2 = 4.602$$

3. The preliminary calculations and final answer:

$$SS_{xx} = 370, SS_{xy} = 22.7333, SS_{yy} = 2.9956$$
$$\hat{\beta}_1 = 0.0614414$$
$$SSE = 1.5988333$$
$$S^2 = 0.228$$
$$S = 0.478$$

4. S = 2.145, so we can say that 95% of the observed HDL values will be within 4.29 points of the predicted value. This means 95% of the time our model will not over or under estimate by more than 4.29 points.