Simple Linear Regression

11.6 Finding r the Coefficient of Correlation

1. Below is a list of heights and weights for a random selection of NFL athletes. Use the data below to find the correlation coefficient r for a linear relationship between height and weight among NFL athletes.
(Note: $SS_{xx} = 76.5455, SS_{yy} = 14034.1818, SS_{xy} = 735.0909$)

Subject	1	2	3	4	5	6	7	8	9	10	11
Height (in)	78	74	76	74	72	76	70	72	70	70	75
Weight	270	270	230	225	220	315	201	188	208	209	224

2. In this problem, we have a list of heights and weights for a random selection of men, women, and children. Use the data below to find the correlation coefficient r for a linear relationship between height and weight. Comparing these results to the results from problem one, what effect does the inclusion of a wider swath of humanity have on the correlation coefficient?
(Note: $SS_{xx} = 947.6364, SS_{yy} = 74628.9091, SS_{xy} = 7672.8182$)

Subject	1	2	3	4	5	6	7	8	9	10	11
Height (in)	59	68	70	67	50	56	71	66	76	85	60
Weight	91	159	112	155	51	84	194	120	298	325	128

3. The goal (per game) average and 2010 salaries for ten professional offensive soccer players is provided below. Use the data to calculate r the correlation coefficient for a linear relationship between goals per game and salary.
(Note: $SS_{xx} = 1.13236, SS_{yy} = 712.9, SS_{xy} = 9.578$)

Subject	1	2	3	4	5	6	7	8	9	10
Goals (per game)	0.2	1.18	0.94	0.50	0.00	0.17	0.39	0.42	0.48	0.60
Salary (in millions)	40	38	32	25	24	21	20	17	17	15

4. A study which looked at marital weight gain (weight gained over the course of a marriage) and marital satisfaction. The correlation coefficient turned out to be -0.904. What does this indicate?

Answers:

1. $r = 0.709$ which indicates a moderate to strong positive linear correlation.

2. $r = 0.912$ which indicates a much stronger positive linear correlation than we had in problem one. The reason for this is that the data set in problem two uses a much wider range of heights and weights. Whenever you use only a narrow span of the possible x-values (or y-values) the other differences that affect the y-values (or x-values) play a more dominant role and dilute the r-value for the relationship you wish to study. By only looking at relatively tall males, we have diluted the strength of the correlation. The degree of correlation between two variables depends on the range of the variables considered. SAT scores/college GPA correlation studies suffer from the same narrowing of the range of the x-values. The SAT scores for admitted students at each school are from a narrow band based on the selectivity of the schools. This means SAT/GPA correlation studies usually do not show better than a 0.30 to 0.45 linear correlation coefficient.

3. $r = 0.337$ which shows weak positive correlation. I think most people would have expected stronger correlation considering the main objective for an offensive player is to score goals, but evidently the relationship between goals and salary is an imperfect one.

4. The correlation coefficient in this problem indicates strong linear correlation between marital weight gain and marital satisfaction. That means that higher weight gain appears with low marital satisfaction and low weight gain appears with high marital satisfaction. Do not make the mistake of assuming causation here. We do not know why these variables

appear together. We cannot assume weight gain causes the marriages to become unhappy. There could be more than one plausible explanation. A third variable acting behind the scenes could be producing the effect we are seeing. For example, what if depression is causing the weight gain and the marital dissatisfaction? In that scenario, it is not the weight gain producing the marital dissatisfaction, but rather the depression produces both the weight gain and the dissatisfaction.