Data Mining – Student Assessments

Assessment – Quiz to accompany Lesson 1

The modern Olympic Games are a modified revival of the Greek Olympian Games that came to be largely through the efforts of the French sportsman and educator Baron Pierre de Coubertin. The Games are an international athletic competition that has been held at a different site every four years since their inauguration in 1896, with occasional interruptions in the times of world wars.

The data for the gold medal performances in long jump, high jump, discus throw are given below (in inches). Source: http://exploringdata.net/oly_gold.htm

years	long jump	high jump	discus throw
-4	249.75	71.25	1147.5
0	282.875	74.8	1418.9
4	289	71	1546.5
8	294.5	75	1610
12	299.25	76	1780
20	281.5	76.25	1759.25
24	293.125	78	1817.125
28	304.75	76.375	1863
32	300.75	77.625	1948.875
36	317.3125	79.9375	1987.375
48	308	78	2078
52	298.8	80.32	2166.85
56	308.25	83.25	2218.5
60	319.75	85	2330
64	317.75	85.75	2401.5
68	350.5	88.25	2550.5
72	324.5	87.75	2535
76	328.5	88.5	2657.4
80	336.25	92.75	2624
84	336.25	93.5	2709.25
88	343.25	93.5	2709.25
92	342.5	92	2563.75

1. Fit the least squares linear regression line using Years as the explanatory variable and Long Jump as the response variable. (3 points)

2. Find the predicted long jump for the year 1996. (2 points)

3. Would this model be acceptable to use to predict the long jump distance for 2008 or 2012? Explain (3 points)

4. The output from the **lm** command and the **cor** command are given below.
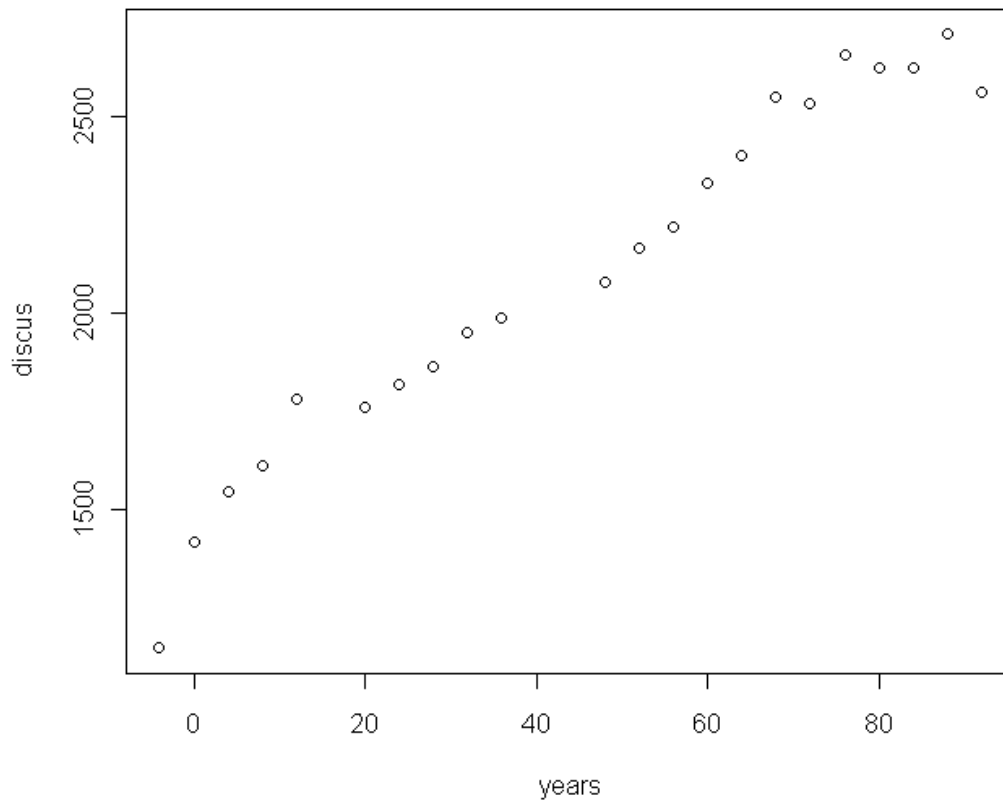
Call:
lm(formula = discus ~ years)

Coefficients:
(Intercept)   years
  1452.19   14.39

> cor(discus, years)
[1] 0.9793683

   a) What is the least squares regression line? Be sure to identify the explanatory variable and
      the response variable. (3 points)
   b) Sketch the regression line on the graph below. (1 points)

   c) The correlation coefficient is 0.9794. Explain what this means in the context of the problem.
      (3 points)
   d) What is the value of the coefficient of determination? What does this value represent in the
      context of this problem? (4 points)