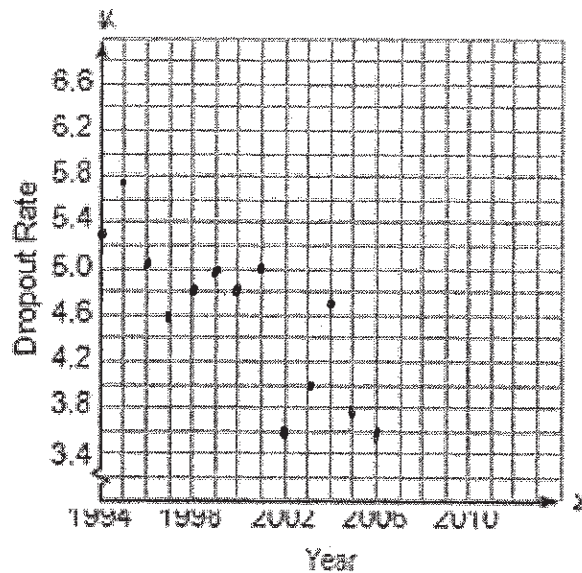


Is It Getting Hot In Here? Modeling Data Using Linear Regression

1. One of the jobs of the National Center for Education Statistics is to gather information about public high schools and their dropout rates. This includes anyone who leaves school without a high school diploma or an equivalent credential. The table below shows the average percent of high school dropouts from October 1994 through October 2006.

Year	High School Dropout Rate
1994	5.3
1995	5.7
1996	5.0
1997	4.6
1998	4.8
1999	5.0
2000	4.8
2001	5.0
2002	3.6
2003	4.0
2004	4.7
2005	3.8
2006	3.6

- a. Create a scatter plot of the high school dropout data. What information can you gather about the dropout rates from the scatter plot?



- b. Determine the linear regression equation for the high school dropout data. Is this line a good representation of the data? Explain your reasoning.

where 1994 is time $x=0$ $y = -.138x + 5.44$

- c. Interpret the slope and y-intercept of the linear regression equation. What do these values represent in terms of the problem situation?

The slope of $-.138$ shows that the high school dropout rate is decreasing by about ~~0.138~~ $.138$ students per year. The y-int of 5.44 says that the high school dropout rate was 5.44% in year $0 \rightarrow 1994$.

- d. Determine the dropout rate for the year 2002. Is this the same as the dropout rate recorded in the table? If not, explain the difference.

2002 is year 8 $\rightarrow y = -.138(8) + 5.44 = 4.336\%$ which is not the same as the value in the table. This is because the line of best fit only gives an estimate based off of the data. The table gives the exact values.

- e. During what year will the dropout rate be 2.5 percent? Show your work and explain your reasoning.

$$2.5 = -.138x + 5.44$$

$$21.3 = x$$

I plugged 2.5 in for y bc that represents the dropout rates. Then I solved for x to find the corresponding year. About 21.3 years after 1994 in year 2015 the dropout rate will be about 2.5.