**Barry Bonds’ Home Runs**

Many consider San Francisco Giants’ slugger Barry Bonds to be one of the greatest baseball players of all time. In 2001, he hit 73 home runs and broke the record for the most home runs hit in a single season. However, in 2011, Bonds was found guilty of obstruction of justice due to his testimony on steroid use which was found to be evasive and misleading. His possible steroid use has led many to question whether his home run record should be allowed to stand.

Barry Bonds played in the major leagues for 22 years. His home run and games played statistics for each season are displayed in the table below.

|  |  |  |
| --- | --- | --- |
| **Season** | **Home Runs** | **Games Played** |
| 1986 | 16 | 113 |
| 1987 | 25 | 150 |
| 1988 | 24 | 144 |
| 1989 | 19 | 159 |
| 1990 | 33 | 151 |
| 1991 | 25 | 153 |
| 1992 | 34 | 140 |
| 1993 | 46 | 159 |
| 1994 | 37 | 112 |
| 1995 | 33 | 144 |
| 1996 | 42 | 158 |
| 1997 | 40 | 159 |
| 1998 | 37 | 156 |
| 1999 | 39 | 102 |
| 2000 | 49 | 143 |
| 2001 | 73 | 153 |
| 2002 | 46 | 143 |
| 2003 | 45 | 130 |
| 2004 | 45 | 149 |
| 2005 | 5 | 14 |
| 2006 | 26 | 130 |
| 2007 | 28 | 126 |

Below is a scatterplot showing the number of home runs Barry Bonds hit each season.



1. From the table and the graph, identify two points that appear to be outliers.
2. Give a possible cause for each outlier.
3. In 2005 Barry Bonds missed most of the season due to a knee injury. Use the table or the scatter plot to explain whether this might have affected his performance in 2006 and 2007.
4. Calculate the regression line for 1986 through 2004, which are the years before his knee injury. (Suggestion: Let the independent variable be years since 1986.)
5. Determine the correlation coefficient *r* for the data from 1986 to 2004.
6. Use your regression line to predict the number of home runs Barry Bonds would hit in 2001? How does this compare with the actual value?
7. Now eliminate the data for the year 2001. Calculate the regression line for 1986 through 2004, excluding the year 2001. (Suggestion: Let the independent variable be years since 1986.)
8. Determine the correlation coefficient *r* for the data from 1986 to 2004, excluding the year 2001.
9. Compare the two regression lines:
   1. Which has the greater slope?
   2. Which shows a stronger correlation?
   3. Explain the differences.
10. If we include all 22 years of Bonds’ career in finding a regression line, how would this affect the slope and the strength of the correlation? Make a prediction and then test it using a calculator or appropriate software.