

AP Statistics: Linear Regression Review – For each of the data sets, perform a complete bivariate analysis by performing all of the following steps:

- a) Make a scatterplot on your calculator and sketch it on your paper.
 - b) Determine the association (pos or neg) and describe it in a sentence.
 - c) Find the equation of the least squares linear regression line.
 - d) Describe the meaning of the slope of the line in a sentence.
 - e) Find the correlation coefficient. What does it tell you about the data?
 - f) What is r^2 and what does it tell you about the data?
 - g) Make a residual plot on your calculator and sketch it on paper. Interpret the plot.
 - h) Are there any outliers/influential points? If so, remove them and recalculate the regression equation. What effect was there on the correlation and regression equation?
-

1. The following are median heights, in inches, of boys age 2 to 14.

Age (yr)	Median Height (in)
2	35.1
3	38.7
4	41.3
5	44.1
6	46.5
7	48.6
8	51.7
9	53.7
10	56.1
11	59.5
12	61.2
13	62.9
14	63.6

2. The following are population (in millions) and number of police officers (in thousands) for various states.

State	Pop	Police
CA	30.4	86.2
CO	3.4	9.2
FL	13.2	45.0
IL	11.5	39.9
IA	2.8	6.0
LA	4.2	11.8
ME	1.2	2.9
MS	5.2	14.6
NJ	7.7	30.5
TN	5.0	12.3
TX	17.3	46.2
VA	6.3	15.2
WA	5.0	10.9

3. The following are data on the number of online educational journals from 91-97.

Year	Journals
91	27
92	36
93	45
94	181
95	306
96	1093
97	2459

AP Statistics: Linear Regression Review – For each of the data sets, perform a complete bivariate analysis by performing all of the following steps:

- i) Make a scatterplot on your calculator and sketch it on your paper.
 - j) Determine the association (pos or neg) and describe it in a sentence.
 - k) Find the equation of the least squares linear regression line.
 - l) Describe the meaning of the slope of the line in a sentence.
 - m) Find the correlation coefficient. What does it tell you about the data?
 - n) What is r^2 and what does it tell you about the data?
 - o) Make a residual plot on your calculator and sketch it on paper. Interpret the plot.
 - p) Are there any outliers/influential points? If so, remove them and recalculate the regression equation. What effect was there on the correlation and regression equation?
-

4. The following are data gathered by anthropologists digging in Mayan ruins. X represents the depth of the soil in centimeters and Y is the percent of the mineral *montmorillonite* in the soil.

Depth	% montmorillonite
40	58
50	34
60	32
70	30
80	28
90	27
100	22

5. The following are data on various fast food hamburgers for various states. X represents the grams of fat in the burger and Y is the number of calories.

Fat	Calories
19	410
31	580
34	590
35	570
39	640
39	680
43	660

6. The following are data on the number of dollars (in millions) that the U.S. federal government spent on mathematics research from 1980 to 1990.

Year	Research \$ (in millions)
80	91
81	118
82	128
83	134
84	151
85	184
86	185
87	205
88	212
89	230
90	245