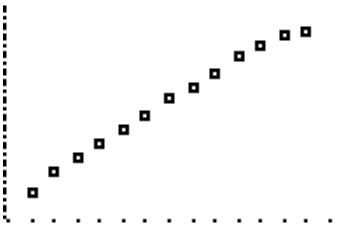
AP Statistics: Linear Regression Review – Answer Key

1. x = age in years, y = median height for boys age 2-14, in inches

   a. Scatter Plot

   b. Positive Association – As age increases, height increases

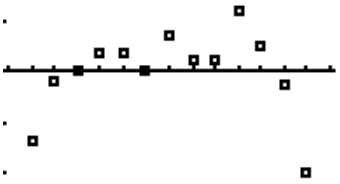
   c. y = 31.567 + 2.429x OR height = 31.567 + 2.429age

   d. On average, boys age 2-14 grow 2.429 inches per year

   e. r = .996. which tells us that there is a strong relationship between age and height

   f. r^2 = 99%, so 99% of the change in height can be attributed to age

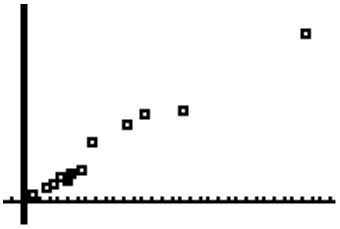
   g. Residual Plot is random, so a line is a good model for data

   h. There are no outliers

2. x = state population in millions, y = number of police officers in 1000's

   a. Scatter Plot

   b. Positive Association – As pop. increases, number of police officers increases

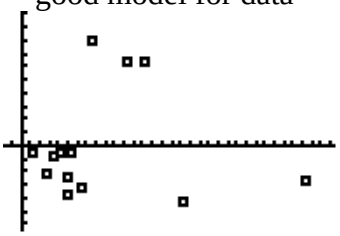
   c. y = -.0008 + 2.921x OR cops = -.0008 + 2.921pop

   d. On average, there are 2,921 police officers per 1 million people

   e. r = .984. which tells us that there is a strong relationship between pop and police

   f. r^2 = 97%, so 97% of the change in police can be attributed to pop

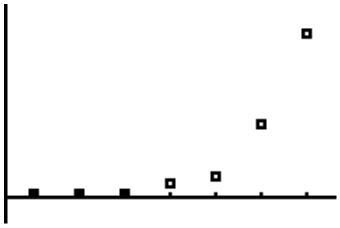
   g. Residual Plot is random, so a line is a good model for data

   h. CA is an outlier. Removal causes r to get lower, equation also changes

3. x = year, y = # of online education journals

   a. Scatter Plot

   b. Positive Association – As time increases, number of journals increases

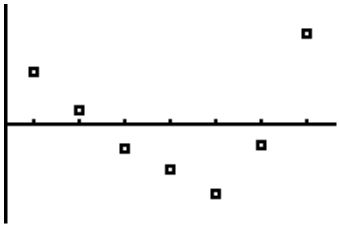
   c. y = -31874.5 + 345.4x OR journals= -31874.5 + 345.4year

   d. On average, the number of journals increases 345.4 per year

   e. r = .824. which tells us that there is a moderate relationship between time and # of journals

   f. r^2 = 68%, so 68% of the change in # of journals can be attributed to time

   g. Residual Plot shows a pattern, so a line is not the best model for data

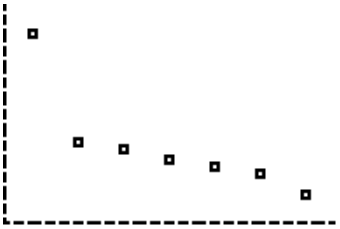
   

   h. There are no outliers

AP Statistics: Linear Regression Review – Answer Key

4. x = soil depth in cm, y = % of *montmorillonite* in the soil

   a. Scatter Plot

   b. Negative Association – As soil depth increases, % *mont*… decreases

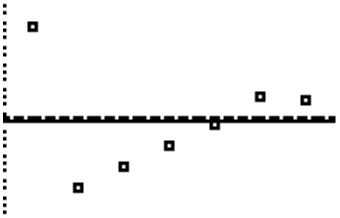
   c. y = 64.5 -.45x OR percent = 64.5 - .45depth

   d. On average, the percent decreases by .45% per cm of soil depth

   e. r = -.833. which tells us that there is a moderate relationship between depth and percent

   f. r^2 = 69%, so 69% of the change in percent can be attributed to soil depth

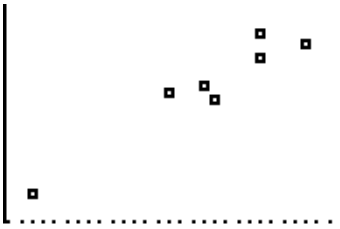
   g. Residual Plot shows a pattern, so a line may not be the best model

   h. (40,58) is an outlier. Removal causes r to get much better (-.98), equation also changes

5. x = fat content in grams, y = calories in burger

   a. Scatter Plot

   b. Positive Association – As fat increases, calories increase

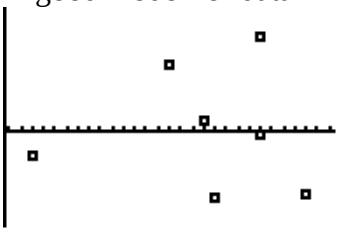
   c. y = 210.95 + 11.056x OR calories = 210.95 + 11.056fat

   d. On average, calories increase 11.056 per gram of fat in burgers

   e. r = .961. which tells us that there is a strong relationship between fat and calories

   f. r^2 = 92%, so 92% of the change in calories can be attributed to fat

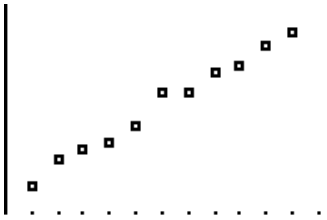
   g. Residual Plot is random, so a line is a good model for data

   h. (19,410) is an outlier. Removal causes r to get lower (.836), equation also changes slightly (the slope)

6. x = year, y = research funding (in millions)

   a. Scatter Plot

   b. Positive Association – As time increases, amount of research funding increases

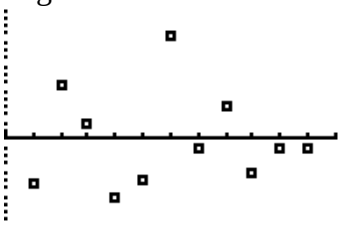
   c. y = -1100.73 + 14.96x OR funding= -1100.73 + 14.96year

   d. On average, the amount of funding increases $14.96 million per year

   e. r = .993. which tells us that there is a very strong relationship between time and research funding

   f. r^2 = 98.6%, so 98.6% of the change in research funding can be attributed to time

   g. Residual Plot is random, so a line is a good model for data

   h. There are no outliers