# MTH 23.5 LECTURE NOTES (Ojakian)

## Topic 13: Correlation and Scatter Diagrams

---

**OUTLINE**

References (**Algebra Book**: None; **Statistics Book**: 12.2)

1. Plotting points
2. Correlation

---

1. How are two variables related?
   (a) Example: Guilded Exercise 1 (ch. 4, p. 122, from 5th edition): Look at just table of numbers.
   (b) Two variables are correlated if: The value of one variable can be used to predict the value of the other variable.
   (c) Goal: Determine how correlated two variables are.

   **PROBLEM 1.** *In the example, guess the work hours lost for various choices of training hours.*

2. Scatter Diagram

   **PROBLEM 2.** *Verify the scatter plot of data for guilded exercise.*

   (a) Terminology
      i. Horizontal axis: Explanatory variable
      ii. Vertical axis: Response variable
      iii. Correlation ...

   (b)

   **PROBLEM 3.** *Make a scatter plot for the following data:*

   $$X: 4, \quad 7, \quad 8, \quad 12, \quad 17$$

   $$Y: 2, \quad 5, \quad 10, \quad 11, \quad 20$$

   *Does the data look "correlated"? What is its rough shape?*

3. Correlation Coefficient

(a) How good is the Best-Fil line? ...
Correlation Coefficient = `Correl`([column 1], [column 2])

(b) Measures how close to a line the scatter plot looks. Denoted $r$.

i. It is between -1 and 1, inclusive.
ii. If $r$ close to 0: Little or no linear correlation.
iii. If $r$ close to +1: Positive correlation
iv. If $r$ close to -1: Negative correlation

(c)

**PROBLEM 4.**

*i. Make up a table of two columns of data, with at least 10 individuals and find the correlation coefficient. Try to choose the data so that $r$ is close to 0.9.*

*ii. Make up a table of two columns of data, with at least 10 individuals and find the correlation coefficient. Try to choose the data so that $r$ is close to $-0.9$.*

*iii. Make up a table of two columns of data, with at least 10 individuals and find the correlation coefficient. Try to choose the data so that $r$ is close to $0$.*

4. Applying Correlation

**PROBLEM 5.** *Pick two variables from class data that you think might be correlated and check.*

5. Correlation versus Causation

"Correlation does not imply causation!"

(a) **Lurking variable (or hidden variable)**: A third variable (not X or Y) that is simultaneously responsible for the changes in X and Y.

(b)

**PROBLEM 6.** *From section 4.1 (5th edition) do problems: 8, 9.*

(c) See webpage: http://www.tylervigen.com/spurious-correlations