**Problem Set #2. MVIS 5301 Statistical Applications for Visualization**

1. Suppose the amount of money earned by all MICA graduates from speaking engagements is normally distributed with a mean of $4,040 and a standard deviation of $510. What percentage of MICA graduates earn between $4,000 and $4,500 on speaking engagements?
2. Find the confidence intervals for the following.
   1. = 20, n = 36, σ = 3, confidence interval = 95%
   2. = 25, n = 36, σ = 3, confidence interval = 95%
   3. = 30, n = 25, σ = 4, confidence interval = 90%
   4. = 50, n = 16, σ = 5, confidence interval = 99%
3. This question asks you to think about the directions of causality. Indicate the possible direction(s) of causality for the relationship between **unemployment** and the **crime rate**. Discuss your reasoning (a few sentences will suffice). Make sure to state clearly the assumptions you are making about the measurement of the construct (i.e., its operationalization) that may be important in justifying your choice.
4. There are 3 parts to this question:
   1. Graph the linear equations and data points
   2. Construct tables for x, y, ŷ, e, and e2;
   3. Determine which line fits the set of data points better, according to the least-squares criterion

Line A: *y* = 1.5 + 0.5*x*

Line B: *y* = 1.125 + 0.375*x*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | 1 | 1 | 5 | 5 |
| y | 1 | 3 | 2 | 4 |

1. An instructor at Arizona State University asked a random sample of eight students to record their study times in a beginning calculus course. She then made a table for total hours studied (*x*) over 2 weeks and the test score (*y*) at the end of the 2 weeks. Here are the results.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 10 | 15 | 12 | 20 | 8 | 16 | 14 | 22 |
| y | 92 | 81 | 84 | 74 | 85 | 80 | 84 | 80 |

The regression equation for these data is: ŷ = 94.86698 – 0.84561*x.*

1. Graph the regression equation and the data points (if you use Excel, don’t simply use the “trendline” feature).
2. Describe the apparent relationship between the two variables under consideration.
3. Interpret the slope of the regression line.
4. What is the predicted exam score for a student who studies for 15 hours?

6. Read this article from Nate Silver at 538 (<http://fivethirtyeight.com/datalab/killing-the-interview-could-cost-sony-100-million/>) and write a paragraph about the potential sources of bias and error in the data and regression methodology.