MAT 340 - CODING ASSIGNMENT #3
due Tuesday, April 16, 2019 at 8:00PM.

OBJECTIVE: Students will design and implement algorithms to compute various statistics. Students will estimate mean, variance and correlation coefficient, use the Central Limit Theorem to compute confidence intervals, and sample uniform distribution to estimate π.

GRADING: The assignment is worth 4% of your course grade.

INSTRUCTIONS:

- Students must work individually and submit their own code, but they may ask questions and clarification from classmates and the instructor.
- Students must submit their project on Moodle.

SUBMIT THE FOLLOWING:

- An executable. This should be able to run on a clean machine, please compile it accordingly (see the MAT 340 forum for a build tip in Visual Studio).
- A data file for problem #3, with student names included in the header.
- A read-me file explaining how to run your code. Include the name of each team member, course, section and semester in the header of your file.

PROJECT:

1. **Estimate** π (9 pts.) You will estimate π by plotting points inside a square and computing the percentage of points that fall inside the inscribed circle. Write a program that does the following:
   - Input the total number of points to plot, N.
   - Generate N pairs of random numbers drawn from the interval $[-1, 1]$.
   - Compute N_c, the number of points that fall inside the unit circle centered at the origin.
   - Use N and N_c to output an estimate for π.
2. **Confidence Intervals** (12 pts.) Write a computer program to simulate 10,000 Bernoulli trials with probability p for success in each trial, compute the average number of successes and estimate the error. Your simulation should do the following:
   - Input p, the probability of success in a Bernoulli trial.
   - In a loop of 100 trials:
     (a) Run 10,000 Bernoulli trials with parameter p and count the proportion of successes.
     (b) Compute the 95% confidence interval for the probability of success, based on the proportion of successes.
     (c) Check if the true value of p is included within the confidence limits.
   - Output the number of times p was in the 95% confidence interval.

3. **Correlation** (17 pts.) For pairs of data (x_i, y_i) that are outcomes of the pairs of dependent random variables X, Y we can estimate the correlation coefficient by

$$\overline{r} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{(n-1) s_X \, s_Y},$$

where $\overline{x}$ and $\overline{y}$ are the sample means for X and Y, and s_X and s_Y are the sample standard deviations for X and Y, respectively. Write a program to compute the sample means, sample standard deviations, and correlation for such dependent data:

- Input data pairs for samples of (X, Y).
- Compute and print the sample mean for X.
- Compute and print the sample mean for Y.
- Compute and print the sample standard deviation for X.
- Compute and print the sample standard deviation for Y.
- Compute and print the correlation coefficient.

The attached spreadsheet provides historical data for one of my courses. You will decide which grade category gives the best predictor for the course grade, by finding the correlation coefficient closest to 1, when running your code for the pairs (Midterm, Course Grade), (Homework, Course Grade), and (Quiz, Course Grade). Print your results to a data file and submit.