Project 2 **Math 111**
Due: Wednesday, 28 March

You maintain a machine for your company that produces USB ports for computers. When the machine was new and working properly, the dimensions of 5000 ports were recorded to be used as a baseline for future comparison. To check the operation of the machine, a random sample of 10 parts is chosen every other day, and the average dimension is recorded for each sample. Here are the observations over a 48-day period.[1]

Day	2	4	6	8	10	12	14	16	18	20	22	24
mean dimension of 10 parts, $\bar{x}$	.4976	.4967	.5010	.5002	.4984	.4967	.4993	.5002	.4976	.5019	.4958	.4993

Day	26	28	30	32	34	36	38	40	42	44	46	48
mean dimension of 10 parts, $\bar{x}$	.4976	.4958	.4984	.5010	.5002	.4993	.5019	.5036	.5028	.5019	.5045	.5054

- Compute the mean and standard deviation (using minitab) of the original 5000 measurement. We will use this baseline mean and standard deviation in place of the mean and standard deviation of the population. Create a histogram of the original data.
- Compute the expected value and standard error of the mean of 10 parts.
- Create a scatterplot showing the day number across the horizontal axis and average port size on the vertical axis. Add horizontal lines (using minitab this time) at the mean, ± 1 SE, ± 2 SE, and ± 3 SE port sizes.

  Example If the mean was .8 and standard error was .1 I could type:

  ```
  plot c3*c2;
  reference 2 .8 .7 .9 .6 1.0 .5 1.1.
  ```

- Save your graphs as images so that you can include them in your report.

Write-up Write a report to your boss (who needs all these statistical terms explained to her) with a recommendation for whether to shut down the machine for maintenance or not. Do not include any minitab code in your report. Be sure to include (at least) the following in your report.

- Your values for the standard error ranges and an explanation of their significance.
- Include your graphs in the relevant section within your report.
- Comment on the patterns of the original data and of the samples taken over the last 48 days.
- If the machine is working properly, the central limit theorem says that the sample means should follow the 68/95/99.7 rule. For the given data, count how many of the sample means are within the limits, and compute the percentage of means within the limits. Does this percentage seem to indicate that the machine is working properly according to the central limit theorem? Explain.
- The central limit theorem only works when the individual samples are large enough. How can we tell here that our samples of size 10 are large enough.
- What is your recommendation for whether to shut the machine down or not? Explain.

[1] A file containing the data is available on the course page of my website, `http://dean.serenevy.net/`.

Extra Credit 1 (Due 2 April)

You are hired by a veterinarian to help improve patient compliance (all shots are gotten on time, heartworm, senior bloodwork, etc.). Below is a table of compliance data for the 12 months prior to your hiring (followed by the compliance data for the month after your hiring). Each number represents the number of animals who have properly complied with the corresponding recommendations of the American Animal Hospital Association.[2]

Month	patients	c2	c3	c4	c5	c6	c7
Apr	621	92	22	279	209	25	123
May	657	87	28	289	197	21	107
Jun	796	100	46	340	262	38	151
Jul	864	99	58	424	259	29	232
Aug	812	135	40	326	259	37	214
Sep	860	169	34	394	249	29	146
Oct	829	114	27	291	254	22	174
Nov	882	127	46	363	276	33	131
Dec	869	125	55	302	262	38	201
Jan	644	90	30	214	171	26	124
Feb	695	124	39	262	225	23	206
Mar	809	78	34	335	254	39	159
May	*933*	*184*	*74*	*428*	*282*	*48*	*292*

You were hired in April, so the May data represent the first month in which the results of your efforts should be visible. Your position is only temporary unless you can show that you are making a real difference in customer compliance. Write a report to the review committee which explains how each of the compliance variables has changed and which variables you can claim have been improved due to your efforts.

Hints/Suggestions:

- Since the number of patients changes each month, each compliance variable should be converted to a percentage before the data are analyzed. For example, `let c12 = c2/c1`
- Use a 97.5% confidence level to decide which variables have varied sufficiently to be unlikely caused by random fluctuations.

Extra Credit 2 (Due 2 April)

Find data on per-child spending for K–12 education by state. Then find average SAT scores by state. Create a scatterplot (using minitab) and compute the correlation coefficient and formula for the the least-squares regression line. Use your regression equations to predict Indiana's average SAT score using Indiana's per-child spending on K–12 education. Similarly predict spending from the average SAT score in Indiana.

Write a letter to the editor or state congress person analyzing how Indiana compares to the national trends. (You do not have to mail your letter, though you may if you wish.) Include a semi-technical supplemental which provides the details of your analysis. Be sure that someone who did not take this course would be able to understand the supplemental! Take into consideration the overall correlation between the variables. Perform the comparison both including and excluding data outliers.

If there is some other topic you are personally interested in feel free to use variables relating to that topic (or even different variables relating to this topic). I merely provide these variables as an example.

[2] A file containing the data is available on the course page of my website, `http://dean.serenevy.net/`.