ENGR 112 Spring 2016
In-Class Activities Matlab 7

This is a team exercise, but you should submit individually. *Your grade for the in-class activities is based on effort.*

Using `normpdf`, `normcdf` and `norminv` for statistical quality control: A sticky wicket

You are a quality control engineer in the manufacturing division of the World Wide Wicket Company (WWWco.) One of your plants has had a higher-than usual number of returns of defective wickets lately, due to the wickets having insufficient wickosity. Based on discussions with the plant managers, you have a suspicion that the wickets produced later in the week are the cause of the plant's problems. You want to determine statistically if this is the case.

You are going to make one long Matlab script for this assignment, adding to it at each step of the way.

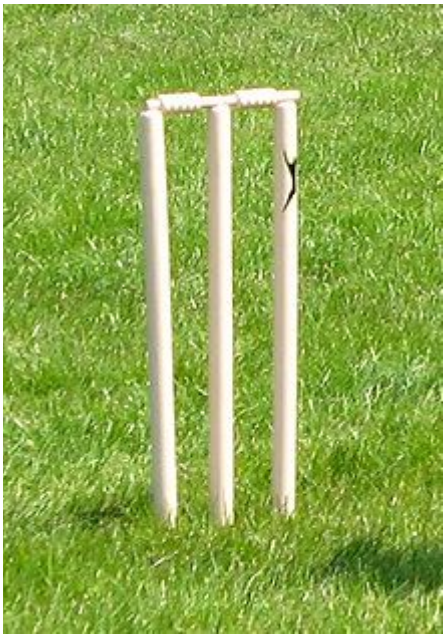


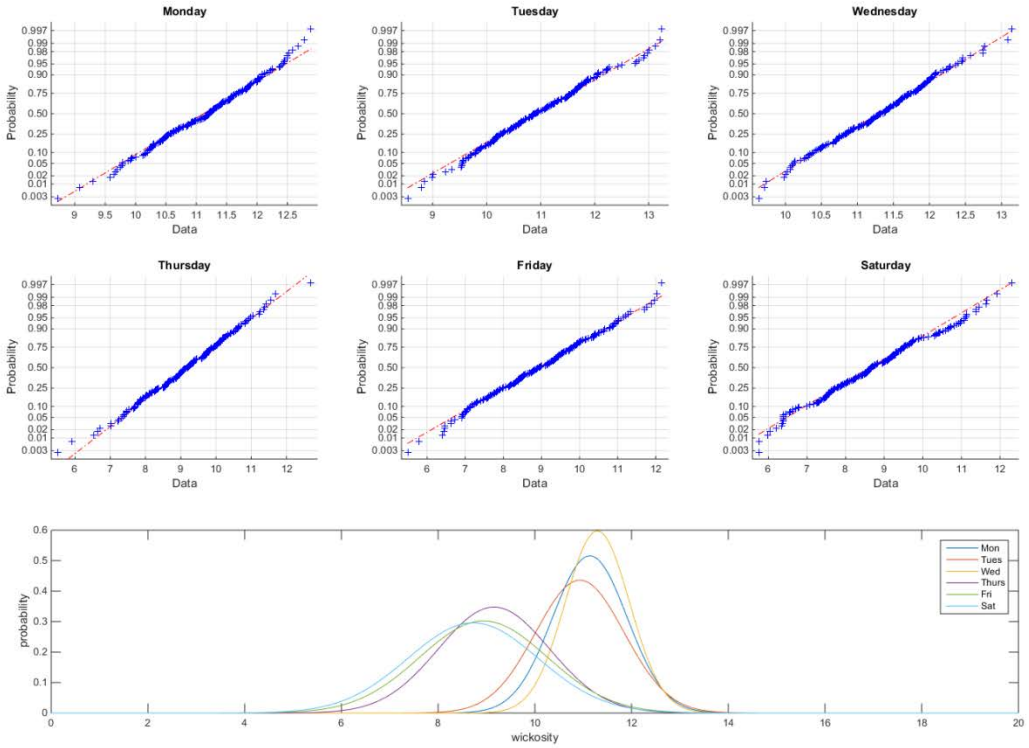
Figure 1: A WWWco wicket

1. For each day of the week, you pull 200 wickets off the production line and measure their wickosity. (The plant produces millions of wickets a day, so your sample of 200 does not affect plant production.) Demand for wickets is so high that the plant runs Monday through Saturday. The data file `wickosity.mat` contains these wickosity measurements. There are 6 columns in the data file, one for each day (except Sunday, when the plant is closed.) There are 200 rows in each column, and each row contains a wickosity measurement.

   First, write a Matlab script to

   - Load the data
   - Use normplot to visually determine if the normal distribution is acceptable for each day's data.
   - Determine the mean and standard deviation of wickosity for each day, and output it neatly in a table.
   - Using these sample means and standard deviations and the function normpdf, make a plot of the estimated pdf for each day, all on the same graph.

   Please use subplot (with a 3 X 3 grid) to put these plots on the same figure. Please be sure to label and title the plots correctly. Your figure should look like this:

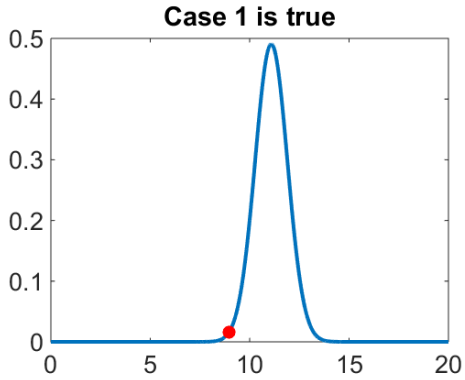


2. It looks to you that while the data from all days may be acceptably normally-distributed, the pdfs of Thursday through Saturday are distinct from the pdfs of Monday through Wednesday. You now need to group the wickosity measurements for the first part of the week and determine the mean and standard deviation of those 600 measurements together, and then do the same for the 600 measurements taken later in the week. Please add to your Matlab script code to
    - manipulate the original data into appropriate arrays
    - calculate the needed means and standard deviations
    - present the results in a clearly to the user

    Your additional output should look like this (your numbers will differ slightly):

```
Mean for Mon-Tues-Wed  is 11.12; standard deviation for Mon-Tues-Wed is 0.80
Mean for Thurs-Fri-Sat is 8.94; standard deviation for Thurs-Fri-Sat is 1.29
```

3. For ease of reference, let's call Mon-Tues-Wed Case 1, and Thurs-Fri-Sat Case 2. We want to calculate the probability of seeing a mean as small or smaller as that from Case 2 in a sample if the real distribution is Case 1



and the probability of seeing a mean as big or bigger from Case 1 if the real distribution is Case 2

We can do this using `normcdf`, with different values of mu and sigma depending on which case we assume is correct. Add code to do this to your script. Your additional results should look like this (your numbers may differ slightly):

```
Probability of seeing a mean of 8.94, given mu = 11.12 and sigma = 1.29 is 0.0032
Probability of seeing a mean of 11.12, given mu = 8.94 and sigma = 0.80 is 0.0455
```

4. Finally, you remember you learned something about the difference in two means in ISEN 314 Statistical Control of Quality. You dig out your old notes and find the formula for the confidence interval for the different in two means:

$$(\mu_1 - \mu_2) = (\overline{x_1} - \overline{x_2}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where μ_1 and μ_2 are the true population means, $\overline{x_1}$ and $\overline{x_2}$ are the sample means, $\sigma_1^2$and σ_2^2 are the variances (you can use the sample variances) and $n_1 = n_2 = 200$.

If the two Cases have the same population mean, then $\mu_1 - \mu_2 = 0$, and we expect zero to fall within the confidence range represented by $(\overline{x_1} - \overline{x_2}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. If the two Cases have different means, then 0 will fall outside this range.

Add code to your script to determine if the two means are different at the 99% confidence level $(\alpha = 0.01)$. Use `norminv` to get your z-value (it isn't 1.96). Your additional results should look something like this (your numbers besides the z-value will be slightly different):

```
z-value for two-sided 99% CI is 2.575829
Difference between means is 1.988239
z-value times square root of sum of standard errors is 0.276120
Means are not equal at 99% confidence
```

5. Assuming you can show the mean wickosity of the wickets changes between the Monday-Tuesday-Wednesday set, and the Thursday-Friday-Saturday, what is happening Wednesday night to cause this change?