

Statistics for Engineers, 4C3/6C3

Assignment 4

Kevin Dunn, dunnkg@mcmaster.ca

Due date: 07 February 2011

Assignment objectives

- Be comfortable using distributions in R.
- Using unpaired and paired tests via confidence intervals.
- Construct and use Shewhart process monitoring charts

Question 1 [1.5]

In the previous assignment you collected the snowfall and temperature data for the HAMILTON A weather station. Here are the data again:

- 1990 to 1999 snowfall: [131.2, 128.0, 130.7, 190.6, 263.4, 138.0, 207.3, 161.5, 78.8, 166.5]
 - 2000 to 2008 snowfall: [170.9, 94.1, 138.0, 166.2, 175.8, 218.4, 56.6, 182.4, 243.2]
 - 1990 to 2000 temperature: [8.6, 8.6, 6.9, 7.1, 7.1, 7.7, 6.9, 7.3, 9.8, 8.8]
 - 2000 to 2008 temperature: [7.6, 8.8, 8.8, 7.3, 7.7, 8.2, 9.1, 8.2, 7.7]
1. Use these data to construct a z -value *and* confidence interval on the assumption that the snowfall in the earlier decade (case A) is statistically the same as in 2000 to 2008 (case B).
 2. Repeat this analysis for the average temperature values.
 3. Do these, admittedly limited, data support the conclusion that many people keep repeating: the amount of snow we receive is less than before and that temperatures have gone up?
 4. In the above analysis you had to pool the variances. There is a formal statistical test, described in the course notes, to verify whether the variances could have come from the same population:

$$F_{\alpha/2, \nu_1, \nu_2} \frac{s_2^2}{s_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{1-\alpha/2, \nu_1, \nu_2} \frac{s_2^2}{s_1^2}$$

where we use $F_{\alpha/2, \nu_1, \nu_2}$ to mean the point along the cumulative F -distribution which has area of $\alpha/2$ using ν_1 degrees of freedom for estimating s_1 and ν_2 degrees of freedom for estimating s_2 . For example, in R, the value of $F_{0.05/2, 10, 20}$ can be found from `qf(0.025, 10, 20)` as 0.2925. The point along the cumulative F -distribution which has area of $1 - \alpha/2$ is denoted as $F_{1-\alpha/2, \nu_1, \nu_2}$, and α is the level of confidence, usually $\alpha = 0.05$ to denote a 95% confidence level.

Confirm that you can pool the variances in both the snowfall and temperature case by verifying the confidence interval contains a value of 1.0.

Note: The equations in the printed course notes you downloaded, or bought from Titles, are wrong; the above version is correct. Please update your printouts.

Question 2 [1.5]

The percentage yield from a batch reactor, and the purity of the feedstock are available as the [Batch yield and purity](#) data set. Assume these data are from phase I operation and calculate the Shewhart chart upper and lower control limits that you would use during phase II. Use a subgroup size of $n = 3$.

1. What is phase I?
2. What is phase II?
3. Show your calculations for the upper and lower control limits for the Shewhart chart on the *yield value*.
4. Show a plot of the Shewhart chart on these phase I data.

Question 3 [2]

You want to evaluate a new raw material (B), but the final product's brittleness, the main quality variable, must be the same as achieved with the current raw material. Manpower and physical constraints prevent you from running a randomized test, and you don't have a suitable database of historical reference data either.

One idea you come up with is to use to your advantage the fact that your production line has three parallel reactors, TK104, TK105, and TK107. They were installed at the same time, they have the same geometry, the same instrumentation, *etc*; you have pretty much thought about every factor that might vary between them, and are confident the 3 reactors are identical.

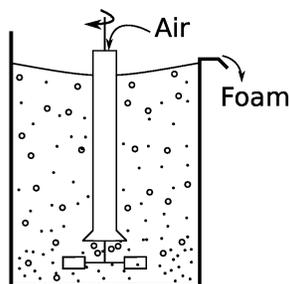
This means that when you do your testing on the new material next week you can run test A using one reactor and test B in another reactor, if you can find the two reactors that have *no statistical difference* in operation.

Normal production splits the same raw material between the 3 reactors. Data [on the website](#) contain the brittleness values from the three reactors for the past few runs using the current raw material (A).

Using a series of paired tests, calculate which two reactors you would pick to run your comparative trial on. Be *very specific and clearly substantiate why* you have chosen your 2 reactors.

Question 4 [2]

A tank uses small air bubbles to keep solid particles in suspension. If too much air is blown into the tank, then excessive foaming and loss of valuable solid product occurs; if too little air is blown into the tank the particles sink and drop out of suspension.



1. Which monitoring chart would you use to ensure the airflow is always near target?
2. Use the [aeration rate dataset](#) from the website and plot the raw data (total litres of air added in a 1 minute period). Are you able to detect any problems?

3. Construct the chart you described in part 1, and show it's performance on all the data. Make any necessary assumptions to construct the chart.
4. At what point in time are you able to detect the problem, using this chart?
5. Construct a Shewhart chart, choosing appropriate data for phase I, and calculate the Shewhart limits. Then use the entire dataset as if it were phase II data.
 - Show this phase II Shewhart chart.
 - Compare the Shewhart chart's performance to the chart in part 3 of this question.

Question 5 [1.5]

Note: For 600-level students

The carbon dioxide measurement is available from a [gas-fired furnace](#). These data are from phase I operation.

1. Calculate the Shewhart chart upper and lower control limits that you would use during phase II with a subgroup size of $n = 6$.
2. Is this a useful monitoring chart? What is going in this data?
3. How can you fix the problem?

Question (not for credit)

Note: This question should take you some time to complete and is open-ended.

A common unit operation in the pharmaceutical area is to uniformly blend powders for tablets. In this question we consider blending an excipient (an inactive magnesium stearate base), a binder, and the active ingredient. The mixing process is tracked using a wireless near infrared (NIR) probe embedded in a V-blender. The mixer is stopped when the NIR spectra stabilize. A new supplier of magnesium stearate is being considered that will save \$ 294,000 per year.

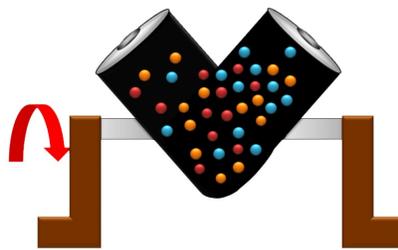


Figure 1: Illustration from Wikipedia (http://en.wikipedia.org/wiki/Industrial_mixer)

The 15 most recent runs with the current magnesium stearate supplier had an average mixing time of 2715 seconds, and a standard deviation of 390 seconds. So far you have run 6 batches from the new supplier, and the average mixing time of these runs is 3115 seconds with a standard deviation of 452 seconds. Your manager is not happy with these results so far - this extra mixing time will actually cost you more money via lost production.

The manager wants to revert back to the original supplier, but is leaving the decision up to you; what would be your advice? Show all calculations and describe any additional assumptions, if required.

END