Statistics for Engineering, 4C3/6C3, 2012

Assignment 3

Kevin Dunn, dunnkg@mcmaster.ca **Due date: 30 January 2012, at noon**

Assignment objectives: working with confidence intervals (calculations and interpretation)

Question 1 [2]

The confidence interval for the population mean takes one of two forms below, depending on whether we know the variance or not. At the 90% confidence level, for a sample size of 13, compare and comment on the upper and lower bounds for the two cases. Assume that $s = \sigma = 3.72$.

$$-c_n \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq c_n$$

$$-c_t \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq c_t$$

Solution

This question aims for you to prove to yourself that the t-distribution is **wider (more broad)** than the normal distribution. The 90% region spanned by the t-distribution with 12 degrees of freedom has upper and lower limits at `qt((1-0.9)/2, df=12)`, i.e. from **-1.782** to **1.782**. The equivalent 90% region spanned by the normal distribution is `qnorm((1-0.9)/2)`, spanning from **z=-1.64** to **z=1.64**. Everything else in the center of the 2 inequalities is the same, so we only need to compare c_t and c_n.

Question 2 [2]

You are responsible for the quality of maple syrup produced at your plant. Historical data show that the standard deviation of the syrup viscosity is 40 cP. How many lab samples of syrup must you measure so that an estimate of the syrup's long-term average viscosity is inside a **range** of 60 cP, 95% of the time.

Solution

We can write the range symbolically as:

$$\text{LB} = \overline{x} - c_n \frac{\sigma}{\sqrt{n}}$$
$$\text{UB} = \overline{x} + c_n \frac{\sigma}{\sqrt{n}}$$

Subtracting and setting equal to 60 cP:

$$\text{UB} - \text{LB} = 60 = 2c_n \cdot \frac{\sigma}{\sqrt{n}}$$
$$n = \left(\frac{(2)(1.96)(40)}{60} \right)^2$$
$$n \approx 7 \text{ samples}$$

Question 3 [4]

The best method of testing for a significant difference is to use an external reference data set. The data I used for the example in class are available on the dataset website, and it only contains the 300 data points from feedback system A.

1. Use these data and repeat for yourself (using any software) the calculations described in class. Reproduce the dot plot, but particularly, the risk value of 11%, from the above data. Note the observations 291 to 300 are the same as 10 "group A" observation in the course slides. The 10 yields from group B are: $[83.5, 78.9, 82.7, 93.2, 86.3, 74.7, 81.6, 92.4, 83.6, 72.4]$.

2. The risk factor of 11% seems too high to reliably recommend system B to your manager. The vendor of the new feedback has given you an opportunity to run 6 more tests, and now you have 16 values in group B:

$$[83.5, 78.9, 82.7, 93.2, 86.3, 74.7, 81.6, 92.4, 83.6, 72.4, \mathbf{70.8}, \mathbf{77.7}, \mathbf{80.7}, \mathbf{81.4}, \mathbf{86.1}, \mathbf{77.9}]$$

   Recalculate the average difference between 2 groups of 16 samples, redraw the dot plot and calculate the new risk factor. Comment on these values and *make a recommendation to your manager*.

   Use bullet points to describe the factors you take into account in your recommendation.

Note: You can construct a dot plot by installing the `BHH2` package in R and using its `dotPlot` function. The `BHH2` name comes from Box, Hunter and Hunter, 2nd edition, and you can read about their case study with the dot plot on page 68 to 72 of their book. The case study in class was based on their example.

Solution

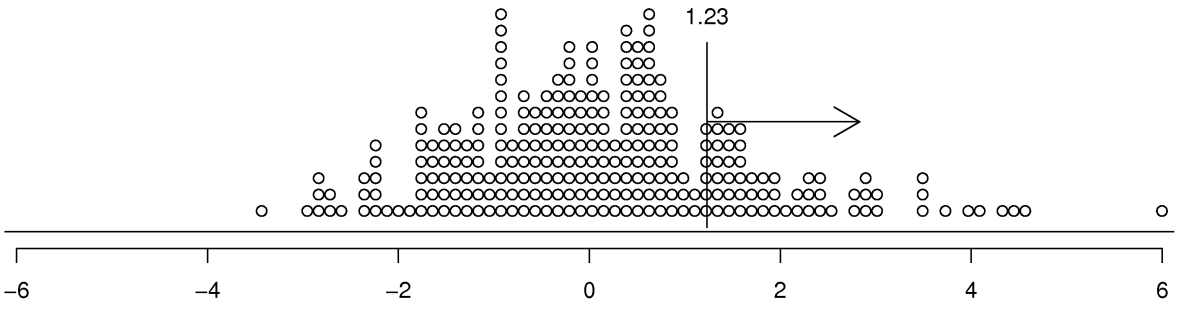
1. You should easily be able to reproduce the dot plot from the notes. Feel free to use Excel, MATLAB, Python or R to calculate the difference between the average of adjacent samples (each of size 10).

2. The key point with calculations in this question are to use adjacent groups, with 16 samples in each group. This is so one can compare like with like. The 16 samples of B have an average of 81.5, while the 16 samples of A immediately prior to B have a mean of 80.3, a seeming improvement of 1.2 percentage points.

   We don't expect this to be a significant improvement at all, because we know the spread in the group A data is large: `sd(yield_A)` from 300 data points is 6.6 units. These 6.6 units are large compared to the average difference of 1.2 units,

   However, the dot plot will show how much of the historical data have a difference of 1.2 units, or greater, between any two adjacent groups containing 16 samples in each group. There are 269 such adjacent groups of 16 samples.

   As the dot plot indicates, about 21.6 percent, or 58 out of 269 groups in the historical data have an average difference of 1.2 or greater. This is an unacceptable risk for most companies: it says that we could be wrong about choosing system B 21% of the time.



This new control system definitely does not have the required improvement from a long-term perspective. The risk is is likely too great for a company to consistently obtain process improvements. The only time control system B might be recommended is for a system where a small percentage increase in improvement will more than pay for the cost of control system B.

Question 4 [2]

You are planning a series of experiments to test alternative conditions in a store and see which conditions lead to higher sales.

Which practical steps would you take to ensure independence in the experimental data, when investigating:

1. adjustable halogen lighting: **A** = soft and dim lighting and **B** = brighter lighting
2. alternative shelving: **A** = solid white metal shelves and **B** = commercial stainless steel racking

Solution

By Cameron DiPietro and Andrew Haines (2012 class)

Randomization is expensive and inconvenient; however, the high cost is to ensure that the results attained in each study are not affected by unmeasured disturbances. We also have to take care to control measured disturbances as far as possible.

1. To ensure independence when investigating adjustable halogen lighting: A = soft and dim lighting and B = brighter lighting, the following experiments and conditions may be run:
    - All light fixtures are changed correctly during the swap from A to B and the same scenario from B to A
    - Keep prices of all products the same during days with A lighting and days with B lighting
    - Do not inform customers of A to B swap or B to A swap in lighting
    - Ensure product quality
    - Use the same amount of voltage throughout the store for each lighting arrangement
    - Keep the store stocked the same for everyday during experiment
    - Use random days for each light fixture
    - Maintain the same advertisements for the store during the study
    - Do not inform employees of lighting swaps to ensure identical employee to customer relationships
    - Compensate for any holiday or unexpected short days of store hours
    - Have employees work randomized shifts to ensure no patterns in employees moods during light fixture swaps
    - Employees have the same mindset to customers (if a retail business) during both A and B lighting arrangements
    - Assume all data from A and B light fixtures have identical population variance

    If lighting A and B are installed simulataneously, then it might be possible to even run different tests during the day, randomly allocated.
2. To ensure independence when investigating alternative shelving: A = solid white metal shelves and B = commercial stainless steel racking, the following experiments and conditions may be run:
    - Shelving size remains the same and in the same location
    - Identical product placement on both shelves A and B, if possible

- Being able to control everything other than the variable being studied of shelves
- Distances between shelves identical
- Ensure employees have the same mindset during each customer visit
- Identical number of items per shelf
- Same shelf distances from checkout
- Clean each shelf in the same manner for both A and B
- Keep prices and sales the same throughout the study period

Clearly the shelf study cannot be easily implemented, since the logistics of unstocking, removing shelf A, replacing with shelf B and restocking them is extremely costly.

One thing to consider in such cases is to run the experiments in two separate stores that are as similar as possible in all other respects (e.g. built in the area with similar profiles of customers, similar store layout, etc.).

Question 5 [3]

There are two analytical techniques for measuring biochemical oxygen demand (BOD). You wish to evaluate the two testing procedures, so that you can select the test which has lower cost, and fastest turn-around time, but without a compromise in accuracy.

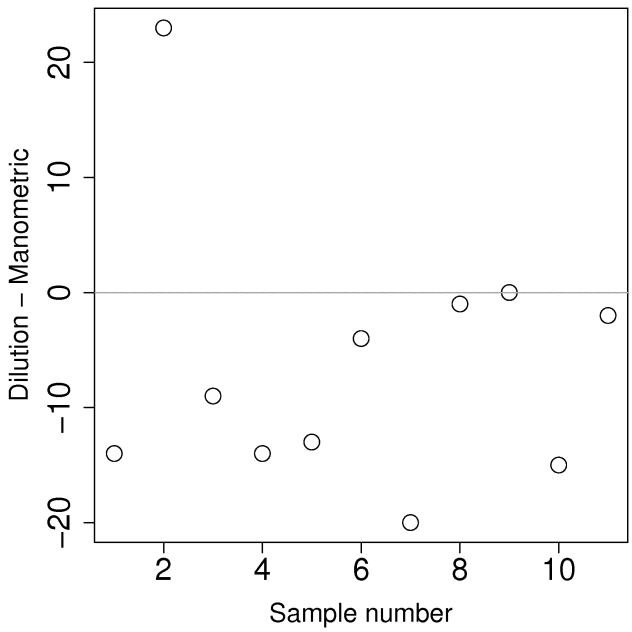
These are the 22 sample values (taken from a single large body of water):

Dilution method	Manometric method
11	25
26	3
18	27
16	30
20	33
12	16
8	28
26	27
12	12
17	32
14	16

1. Is there a statistical difference in accuracy between the two methods?
2. Any other thoughts on the results from the statistical test?

Solution

The temptation is to jump into the code and calculate the t-values and averages differences ($\overline{x}_D = 16.4$, and $\overline{x}_M = 22.6$). But start with a plot of the data, specifically a plot of the differences between the two methods. The immediate problem you see is that average difference of 6.2 between the methods is strongly influenced by a single observation (the second one). In general, the dilution method always produced a smaller result than the manometric method. We expect to see that in our analytical results.



Now let's look at the analytical answer. As before, we can calculate $z = 1.86 = \frac{6.27}{3.375}$ (where $s_p^2 = 62.7$), with a probability of 96.1% that we will have a value smaller than this (risk = 3.9% that we are wrong). A confidence interval would be $-0.77 < \mu_M - \mu_D < 13.3$. And it is at this point that you should realize the problem, even if you didn't plot your data. The fact that the confidence interval only just includes zero is what should raise concern; if the two methods were roughly equivalent, then the interval should span zero with rough symmetry. But this is too close.

So omitting the second point and repeating the analysis gives: calculate $z = 3.24 = \frac{9.20}{2.84}$ (where $s_p^2 = 40.4$), with a probability of 99.8% that we will have a value smaller than this (risk = 0.2% that we are wrong). A confidence interval would be $3.2 < \mu_M - \mu_D < 15.2$; this is a result that is much more aligned with the plotted data.

Note: You may have discovered/used the `t.test(...)` function in R. If you know what you are doing with this function, you are welcome to use it; however I'm reluctant to advocate its use at this point, because these exercises are all about understanding what is going on with confidence intervals and calculating them yourself.

Question 6 [600-level students] [3]

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.

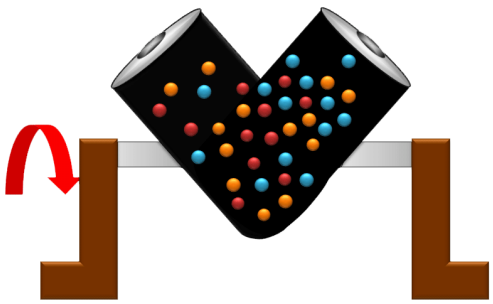


Illustration from Wikipedia (http://en.wikipedia.org/wiki/Industrial_mixer)

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.

Solution

This question, similar to most real statistical problems, is open-ended. This problem considers whether a significant difference has occurred. And in many cases, even though there is significant difference, it has to be weighed up whether there is a *practical* difference as well, together with the potential of saving money (increased profit).

You should always state any assumptions you make, compute a confidence interval for the difference and interpret it.

The decision is one of whether the new material leads to a significant difference in the mixing time. It is desirable, from a production point of view, that the new mixing time is shorter, or at least the same. Some notation:

$$\begin{aligned} \hat{\mu}_{\text{Before}} = \overline{x}_B &= 2715 & \qquad \hat{\mu}_{\text{After}} = \overline{x}_A &= 3115 \\ \hat{\sigma}_{\text{Before}} = s_B &= 390 & \qquad \hat{\sigma}_{\text{After}} = s_A &= 452 \\ n_B &= 15 & \qquad n_A &= 6 \end{aligned}$$

Assumptions required to compare the two groups:

- The individual samples within each group were taken independently, so that we can invoke the central limit theorem and assume these means and standard deviation are normal distributed.
- Assume the individual samples within each group are from a normal distribution as well.
- Assume that we can pool the variances, i.e. σ_{Before} and σ_{After} are from comparable distributions.
- Using the pooled variance implies that the z-value follows the t-distribution.
- The mean of each group (before and after) is independent of the other (very likely true).
- No other factors were changed, other than the raw material (we can only hope, though in practice this is often not true, and a paired test would eliminate any differences like this).

Calculating the pooled variance:

$$\begin{aligned} s_P^2 &= \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A - 1 + n_B - 1} \\ &= \frac{(6-1)452^2 + (15-1)390^2}{6 - 1 + 15 - 1} \\ &= 165837 \end{aligned}$$

Computing the z-value for this difference:

$$\begin{aligned} z &= \frac{(\overline{x}_B - \overline{x}_A) - (\mu_B - \mu_A)}{\sqrt{s_P^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \\ z &= \frac{(2715 - 3115) - (\mu_B - \mu_A)}{\sqrt{165837\left(\frac{1}{6} + \frac{1}{15}\right)}} \\ z &= \frac{-400 - (\mu_B - \mu_A)}{196.7} = -2.03 \qquad \text{on the hypothesis that} \qquad \mu_B = \mu_A \end{aligned}$$

The probability of obtaining this value of z can be found using the t-distribution at 6 + 15 - 2 = 19 degrees of freedom (because the standard deviation is an estimate, not a population value). Using tables, a value of 0.025, or 2.5% is found (in R, it would be `pt(-2.03, df=19) = 0.0283`, or 2.83%). At this point one can argue either way that the new excipient leads to longer times, though I would be inclined to say that this probability is too small to be due to chance alone. Therefore there is a significant difference, and we should revert back to the previous excipient. Factors such as operators, and other process conditions could have affected the 6 new runs.

Alternatively, and this is the way I prefer to look at these sort of questions, is to create a confidence interval. At the 95% level, the value of c_t in the equation below, using 19 degrees of freedom is `qt(0.975, df=19) = 2.09` (any value close to this from the tables is acceptable):

$$\begin{array}{rcccl}
-c_t & \leq & z & \leq & +c_t \\
(\overline{x}_B - \overline{x}_A) - c_t\sqrt{s_P^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} & \leq & \mu_B - \mu_A & \leq & (\overline{x}_B - \overline{x}_A) + c_t\sqrt{s_P^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} \\
-400 - 2.09\sqrt{165837\left(\frac{1}{6}+\frac{1}{15}\right)} & \leq & \mu_B - \mu_A & \leq & -400 + 2.09\sqrt{165837\left(\frac{1}{6}+\frac{1}{15}\right)} \\
-400 - 412 & \leq & \mu_B - \mu_A & \leq & -400 + 412 \\
-812 & \leq & \mu_B - \mu_A & \leq & 12
\end{array}$$

The interpretation of this confidence interval is that there is no difference between the current and new magnesium stearate excipient. The immediate response to your manager could be "*keep using the new excipient*".

However, the confidence interval's asymmetry should give you pause, certainly from a practical point of view (this is why I prefer the confidence interval - you get a better interpretation of the result). The 12 seconds by which it overlaps zero is so short when compared to average mixing times of around 3000 seconds, with standard deviations of 400 seconds. The practical recommendation is that the new excipient has longer mixing times, so "*revert to using the previous excipient*".

One other aspect of this problem that might bother you is the low number of runs (batches) used. Let's take a look at how sensitive the confidence interval is to that. Assume that we perform one extra run with the new excipient ($n_A = 7$ now), and assume the pooled variance, $s_p^2 = 165837$ remains the same with this new run. The new confidence interval is:

$$\begin{array}{rcccl}
(\overline{x}_B - \overline{x}_A) - c_t\sqrt{s_P^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} & \leq & \mu_B - \mu_A & \leq & (\overline{x}_B - \overline{x}_A) + c_t\sqrt{s_P^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} \\
(\overline{x}_B - \overline{x}_A) - 2.09\sqrt{165837\left(\frac{1}{7}+\frac{1}{15}\right)} & \leq & \mu_B - \mu_A & \leq & (\overline{x}_B - \overline{x}_A) + 2.09\sqrt{165837\left(\frac{1}{7}+\frac{1}{15}\right)} \\
(\overline{x}_B - \overline{x}_A) - 390 & \leq & \mu_B - \mu_A & \leq & (\overline{x}_B - \overline{x}_A) + 390
\end{array}$$

So comparing this ± 390 with 7 runs, to the ± 412 with 6 runs, shows that the confidence interval shrinks in quite a bit, much more than the 12 second overlap of zero. Of course we don't know what the new $\overline{x}_B - \overline{x}_A$ will be with 7 runs, so my recommendation would be to perform at least one more run with the new excipient, but I suspect that the new run would show there to be a significant difference, and statistically confirm that we should "*revert to using the previous excipient*".

END