

Case Study – Call Centre Hypothesis Testing

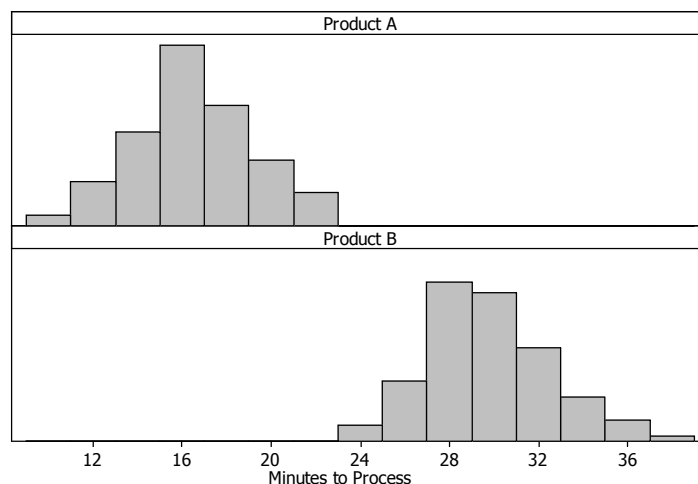
Hypothesis Testing is often thought of as an advanced Six Sigma tool but it is a very useful technique with many applications and in many cases it can be quite simple to use.

Hypothesis tests are used to make comparisons such as –

- Is the new machine faster than the old machine?
- Is the level of complaints higher now than it was last year?
- Is Team 1's quality better than Team 2's?
- Do we receive more queries from older customers than younger customers?
- Are children who take Omega 3 supplements brainier than those who don't?



If there is a large difference then simple graphical analysis can answer the question. In this example, a financial services company wanted to know if the time it took to set up a new account on two different product variants was similar. They collected some cycle time data and plotted it on a histogram. Clearly, Product B took a lot longer.

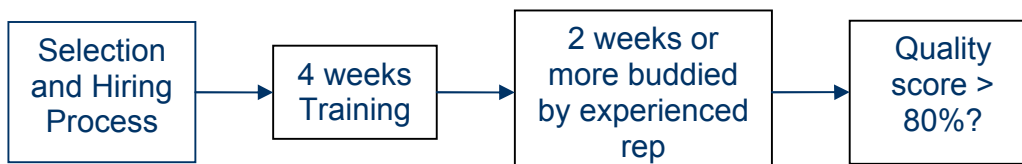


But often when we are faced with comparisons like this, the situation is not so clear cut. This is illustrated in the case study below.

Case Study – Call Centre Hypothesis Testing

Case Study

A Green Belt project in an insurance company call centre was goaled with improving the quality of training for new-hires.



To be signed-off to work alone, the new-hire had to attain an overall quality score (via the standard QC process) of 80%. This was first assessed during their third week after training.

The project identified a number of issues with the original training programme which included –

- Redundant material taught (related to obsolete products)
- No prioritisation of material taught – equal emphasis was placed on topics which were rarely needed
- Complex on-line help system

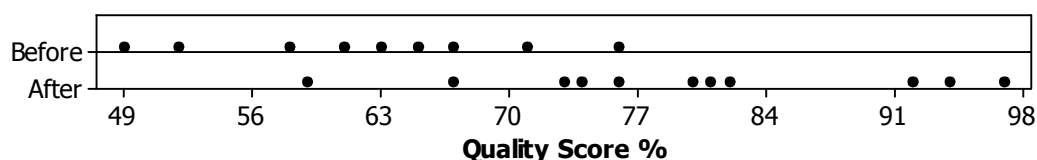
Solutions were designed for these issues. The training was redesigned and could now be delivered in 16 days. The various help systems were combined into one.

The quality score at Week 3 assessment of the first group of 11 new hires trained was compared to the last group of 9 who went through the old training process. The groups had both been rated as “typical” when undergoing selection.

| Before | After | |
|--------|-------|--------------------|
| 58 | 74 | Quality Scores (%) |
| 52 | 97 | |
| 71 | 82 | |
| 63 | 73 | |
| 76 | 80 | |
| 67 | 94 | |
| 49 | 81 | |
| 61 | 76 | |
| 65 | 59 | |
| | 67 | |
| | 92 | |

Case Study – Call Centre Hypothesis Testing

The data was analysed using Minitab. Firstly, a stratified frequency plot called a dot plot was produced. The dot plot gives a clearer picture than a histogram when the sample size is small (as it is here).



It does appear that the new process is producing higher scores. But if we took two other, similar groups of new-hires and put them through the each process, would we get exactly the same scores as we did this time?

The answer is no. We might get similar results but it's very unlikely they will be exactly the same. It's like if you timed your journey to work every day with a stopwatch; the journey time would always vary because of differences in traffic volumes, how many lights you had to stop at, weather conditions and so on. These reasons are called "common causes" of variation and there are always lots of them affecting all processes. In the case of the new hire training, some common causes would have included the particular calls which came in, how well the systems were working, motivation levels etc.

That seems to leave us in a somewhat uncertain position. However, what a hypothesis test does, is look at the size of the difference between the two things (Before and After in this case) and compare this to the size of the common cause variation. If it is bigger than common cause variation, then we can say with confidence that there is a real difference (a "statistically significant" difference).

The type of hypothesis test used here is called a t-test which is used for comparing the averages of two groups. Minitab's output is shown below.

Case Study – Call Centre Hypothesis Testing

Two-sample T for Before vs After

| | N | Mean | StDev | SE Mean |
|--------|----|-------|-------|---------|
| Before | 9 | 62.44 | 8.63 | 2.9 |
| After | 11 | 79.5 | 11.6 | 3.5 |

Difference = mu (Before) - mu (After)

Estimate for difference: -17.1010

95% CI for difference: (-26.6507, -7.5513)

T-Test of difference = 0 (vs not =): T-Value = -3.78 P-Value = 0.002 DF = 17

- ❶ Shows the mean (average) score of each of each group. The difference between the means is 17.1%
- ❷ The t-test outcome is a p-value of 0.002. By convention, the interpretation is that a p-value less than 0.05 means there is a statistically significant difference between the two groups. We can therefore safely conclude that the After process is producing a higher average score - a difference of 17% in means is very unlikely to be the result of common cause variation. Or put yet another way, 17% is bigger than the common cause variation.
- ❸ Going a little deeper, 17% is what we call a "point estimate" for the difference between the two groups. Because of common cause variation, sometimes it will be 17%, sometimes more than 17% and sometime less. A useful way to be able to state this is that the 95% confidence interval runs from 7% to 26%. In other words, we expect the true difference between Before and After to lie between 7% and 26%.

Case Study Conclusion

Using the hypothesis test, the team were able to present strong evidence that the new method was better, producing an improvement in Week 3 score of between 7% and 26%. This would result in more new hires being signed-off and becoming productive more quickly. The new method was adopted and a control plan implemented to verify that future performance was in line with what the trial predicted.

Case Study – Call Centre Hypothesis Testing



Final Comments

Whenever improvements or other differences need to be established, hypothesis tests provide the strong, fact-based, evidence as to whether the difference is real or is possibly just apparent (due to common cause variation). This means that business decisions can be based on fact. The cost of implementing the change can be balanced with the expected size of the benefit. Without the hypothesis test, the cost of the change can only be balanced with a guess.