2-Sample t-Test for Drinking Water Quality

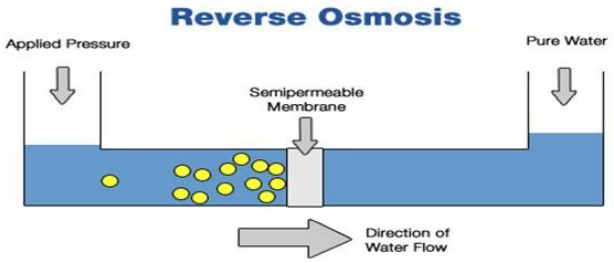
"Ari" Arivazhagan, PMP, PMI-SP,PMI-RMP,MBB

Founder-CEO at International Institute of Project Management (IIPM)

Business Case : I was referred recently, by a large housing complex resident- owner, to resolve a dispute between two drinking water suppliers. The dispute was that both the suppliers claimed their water quality was better than the other. The dispute was resolved using Business Analytics tools to analyze the sample data and arrive at a more informed quantitative decision that would be acceptable to all the stakeholders of the dispute; the association office-bearers, the consumers, and the water suppliers.

The Scenario : KSM Water Suppliers, a drinking and utility (raw) water supplies company, supplies water to a large housing complex (with about 400 families) that purchases water from external sources for domestic drinking /utility purposes through its own fleet of water-tankers. A dispute was raised by a few residents saying that the water supplied by KSM was of very poor quality and is not usable for drinking/utility purposes (washing, cleaning, bathroom use etc). The Secretary of the Housing Complex Association identified another water supplier (Supplier-2-Comp) who supplies, they claimed, better quality of water than the previous supplier (KSM). How would Supplier-1 justify that his water is of better quality than the new supplier in order to retain the contract to supply water regularly to such a huge housing complex?

The Analytics Premise : It was recommended to do a 2-Sample t-test on the samples of water from both suppliers to see if there was any significant difference. Water samples from both the supplier's were collected and two residents were picked randomly and were asked to check the quality of water using TDS meters for 15 samples collected from each Supplier.



To analyze it, first we need to construct the Null Hypothesis (Ho) and the Alternate Hypothesis (Ha) for the claim. Then, let us say we want to be 95% confident in declaring our results. This means, we are willing to take a chance of 5% error in our judgment. If this is OK with us, then the significance level will be 5% (Alpha,$\alpha = 0.05$) and the Confidence Level will be 95%.

Ho : There is no difference between the water samples of Supplier-1 & Supplier-2

Ha : There is difference between the water samples of Supplier-1 & Supplier-2

The Measurement System : Two (2) residents from the housing colony were selected randomly and each of them was asked to check the water quality from Supplier-1 and Supplier-2. Each

resident measured the value of Total Dissolved Solids (TDS) 3 times (3 trials) using digital TDS meters. Average of 3 trials were recorded as one measured value of TDS for one sample.The digital TDS meter used is shown in Figure 1. In total, 15 random samples (each with 3 trials) for each Supplier-1 & Supplier- 2 were collected and the data of two independent water suppliers (as shown in Table -1) were analyzed using Minitab® to see if there was significant difference between the water qualities.

Figure-1 Digital TDS Meter (HM).

C5	C6
Supplier- 1 (KSM)	Supplier-2 (Comp)
613	654
616	655
603	652
607	660
608	655
621	657
610	659
616	658
621	660
611	659
613	652
611	657
607	660
621	660
616	654

Table-1 : TDS values in ppm

```
Two-Sample T-Test and CI: Supplier- 1 (KSM), Supplier-2 (Comp)

Two-sample T for Supplier- 1 (KSM) vs Supplier-2 (Comp)

                    N    Mean  StDev  SE Mean
Supplier- 1 (KSM)  15  612.93   5.56      1.4
Supplier-2 (Comp)  15  656.80   2.93     0.76


Difference = mu (Supplier- 1 (KSM)) - mu (Supplier-2 (Comp))
Estimate for difference:  -43.87
95% CI for difference:  (-47.24, -40.49)
T-Test of difference = 0 (vs not =): T-Value = -27.02  P-Value = 0.000  DF = 21
```

Figure- 3 : Results of 2-Sample t-test from Minitab®

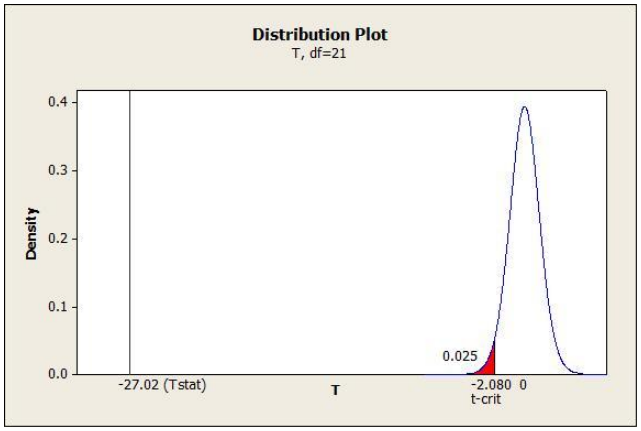


Figure-4 Output from Minitab® t-distribution

Inference:

Referring to the 2-Sample t-test analysis results shown above, in Figure 3, and since the P-Value = 0.000 < 0.05, we reject the Null Hypothesis (Ho). That means, we accept (i.e fail to reject) the alternate hypothesis (Ha) that there is a significant difference between water samples of Supplier-1 and Supplier-2. We can make this statement with 100% confidence since the P value = 0.000. According to the World Health Organization (WHO), the acceptable contamination level (TDS) for drinking water is <= 500 mg / liter (i.e. 500 ppm). Please note that both suppliers' water is not fit for drinking purposes as their ppm level is greater than the permitted ppm level for potable/ drinking water by WHO without further treatment. However, as we were told later that the suppliers'

waters were also used for utility purposes, the water supplied by both the suppliers are acceptable to the purpose intended with the different TDS levels measured. From P-Value criterion, it can be concluded that Supplier-1 (KSM) water is significantly different and better than the Supplier-2 (Comp) based on the P-Value and the TDS levels.

Note : (Updated on 05.01.2015 based on a Reader's query) Although, for convenience sake, it's a common practice to compare the P-value with the Alpha vlaue($\alpha = 0.05$) to decide about the acceptance or rejection of hypotheses, in a t-test it is a good practice to reject (or fail to reject) null hypothesis based on the comparison of Tstat and the critical t-value (for $\alpha = 0.05$, DF=21) . In this case study, from Figure-4,since the Tstat (-27.02) is less than the tcrit (-2.08), we reject the null hypothesis as we did before based on the P-Value criteria.

Further Works : Interested candidates can pursue further on this case by taking more samples on different days in a month to increase the power of the test with various significance levels.

Other Applications: 2-Sample t-test is very useful to test or solve many common business problems / claims by companies to validate if their claims are valid and helps solve business disputes. This test can be used to answer questions like whether the average success rate is higher after implementing a new sales tool than before or whether the test results of patients who received a drug are better than test results of those who received a placebo etc.

Research Assistant : R.Krithika

Data Analyst : A.Karthik

References :

1. Water Network Research

2. **Minitab®** Software

Picture Courtesy : www.officedesignideas.com
Digital TDS meter : www.hmdigital.com

About the Author: The author, Ondiappan Arivazhagan "Ari", is an Honors graduate in Civil / Structural Engineering from University of Madras.He is a certified PMP, PMI-SP, PMI-RMP from PMI, USA. He is also a Master Black Belt in Lean Six Sigma and has done Business Analytics from IIM, Bangalore. He has 30 years of professional global project management experience in various countries around the World and has almost 14 years of teaching / training experience in Project management, Analytics, Risk Management and Lean Six Sigma .He is the Founder-CEO of International Institute of Project Management (IIPM), Chennai and can be reached at askari@iipmchennai.com