Making Sense of the Two-Sample T-Test

font size | Print | Be the first to comment!

Rate this item   (1 vote)

The two-sample t-test is one of the most commonly used hypothesis tests in Six Sigma work. It is applied to compare whether the average difference between two groups is really significant or if it is due instead to random chance.

By Chew Jian Chieh

The two-sample t-test is one of the most commonly used hypothesis tests in Six Sigma work. It is applied to compare whether the average difference between two groups is really significant or if it is due instead to random chance. It helps 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.

Here is an example starting with the absolute basics of the two-sample t-test. The question being answered is whether there is a significant (or only random) difference in the average cycle time to deliver a pizza from Pizza Company A vs. Pizza Company B. This is the data collected from a sample of deliveries of Company A and Company B.

Table 1: Pizza Company A vs. Pizza Company B Sample Deliveries	
A	B
20.4	20.2
24.2	16.9
15.4	18.5
21.4	17.3
20.2	20.5
18.5	
21.5	

To perform this test, both samples must be normally distributed.

Figure 1: Probability Plot of Sample A and Sample B
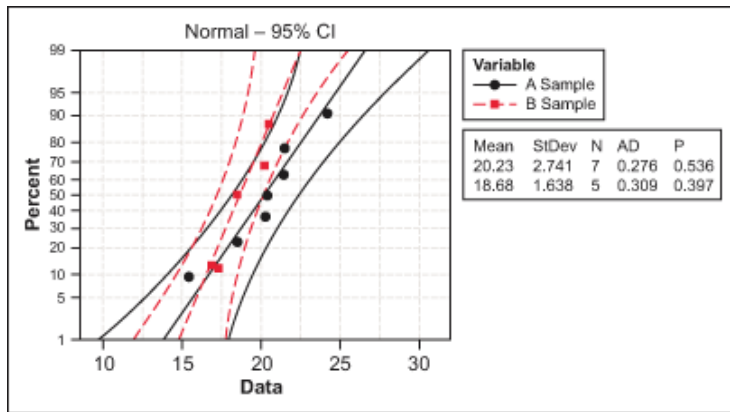
Since both samples have a p-value above 0.05 (or 5 percent) it can be concluded that both samples are normally distributed. The test for normality is here performed via the Anderson Darling test for which the null hypothesis is "Data are normally distributed" and the alternative hypothesis is "Data are not normally distributed."

Using the two-sample t-test, statistics software generates the output in Table 2.

Table 2: Two-Sample T-Test and Confidence Interval for A Sample and B Sample				
	N	Mean	Standard Deviation	SE Mean
A Sample	7	20.23	2.74	1.0
B Sample	5	18.68	1.64	0.73

Difference = mu (A Sample) - mu (B Sample)
Estimate for difference: 1.54857
95% CI for difference: (-1.53393, 4.63107)
T-test of difference = 0 (vs not =): T-value = 1.12, P-value = 0.289, DF = 10
Both use pooled StDev = 2.3627

Since the p-value is 0.289, i.e. greater than 0.05 (or 5 percent), it can be concluded that there is no difference between the means. To say that there is a difference is taking a 28.9 percent risk of being wrong.

If the two-sample t-test is being used as a tool for practical Six Sigma use, that is enough to know. The rest of the article, however, discusses understanding the two-sample t-test, which is easy to use but not so easy to understand.

So How Does It Work?

Actually, if one subtracts the means from two samples, in most cases, there will be a difference. So the real question is not really whether the sample means are the same or different. The correct question is whether the population means are the same - i.e., are the two samples coming from the same or different populations?

Hence, $\overline{X}_1 - \overline{X}_2$ will most often be unequal to zero.

However, if the population means are the same, $\mu_1 - \mu_2$ will equal zero. The trouble is, only two samples exist.

The question that must be answered is whether $\mu_1 - \mu_2$ is zero or not.

The first step is to understand how the one-sample t-test works. Knowing this helps to answer questions like in the following example: A supplier of a part to a large organization claims that the mean weight of this part is 90 grams. The organization took a small sample of 20 parts and found that the mean score is 84 grams and standard deviation is 11. Could this sample originate from a population of mean = 90 grams?

Figure 2: Equation Values

$$\mu_{H_o} = 90$$
$$n = 20$$
$$\overline{X}_1 = 84$$
$$s = 11$$

The organization wants to test this at significance level of 0.05, i.e., it is willing to take only a 5 percent risk of being wrong when it says the sample is not from the population. Therefore:

Null Hypothesis (H0): "True Population Mean Score is 90"
Alternative Hypothesis (Ha): "True Population Mean Score is not 90"
Alpha is 0.05

Logically, the farther away the observed or measured sample mean is from the hypothesized mean, the lower the probability (i.e., the p-value) that the null hypothesis is true. However, what is far enough? In this example, the difference between the sample mean and the hypothesized population mean is 6. Is that difference big enough to reject H0? In order to answer the question, the sample mean needs to be standardized and the so-called t-statistics or t-value need to be calculated with this formula:

$$t = \frac{\overline{X} - \mu_{H_o}}{\hat{\sigma}_{\overline{X}}}$$

In this formula, $\hat{\sigma}_{\overline{X}}$ is the standard error of the mean (SE mean). Because the population standard deviation is not known, we have to estimate the SE mean. It can be estimated by the following equation:

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

where $\hat{\sigma}$ is the sample standard deviation or s.

In our example, $\hat{\sigma}$ is:

$$\hat{\sigma}_{\overline{X}} = \frac{11}{\sqrt{20}} = 2.46$$

Next we obtain the t-value for this sample mean:

$$t = \frac{\overline{X} - \mu_{H_o}}{\hat{\sigma}_{\overline{X}}} = \frac{84 - 90}{2.46} = -2.44$$

Finally, this t-value must be compared with the critical value of t. The critical t-value marks the threshold that – if it is exceeded – leads to the conclusion that the difference between the observed sample mean and the hypothesized population mean is large enough to reject H0. The critical t-value equals the value whose probability of occurrence is less or equal to 5 percent. From the t-distribution tables, one can find that the critical value of t is +/- 2.093.

Figure 5: Finding Critical Value
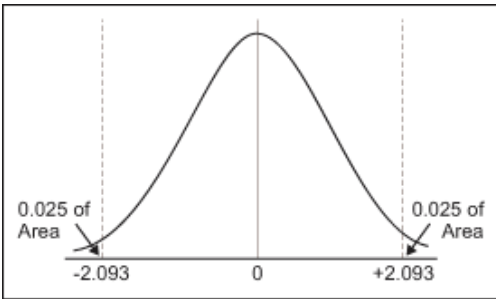
Since the retrieved t-value of -2.44 is smaller than the critical value of -2.093, the null hypothesis must be rejected (i.e., the sample mean is not from the hypothesized population) and the supplier's claims must be questioned.

The Two-Sample T-Test Works in the Same Way

In the two-sample t-test, two sample means are compared to discover whether they come from the same population (meaning there is no difference between the two population means). Now, because the question is whether two populations are actually one and the same, the first step is to obtain the SE mean from the sampling distribution of the difference between two sample means. Again, since the population standard deviations of both of the two populations are unknown, the standard error of the two sample means must be estimated.

In the one-sample t-test, the SE mean was computed as such:

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

Hence:

$$\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n1} + \frac{\hat{\sigma}_2^2}{n2}}$$

However, this is only appropriate when samples are large (both greater than 30). Where samples are smaller, use the following method:

$$\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = Sp\sqrt{\frac{1}{n1} + \frac{1}{n2}}$$

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Sp is a pooled estimate of the common population standard deviation. Hence, in this method it can be assumed that variances are equal for both populations. If it cannot be assumed, it cannot be used. (Statistical software can handle unequal variances for the two-sample t-test module, but the actual calculations are complex and beyond the scope of this article).

The two-sample t-test is illustrated with this example:

Table 3: Illustration of Two-Sample T-Test			
	N	Mean	Standard Deviation

| A Sample | 12 | 92 | 15 |
| B Sample | 15 | 84 | 19 |

Ho is: "The population means are the same, i.e., $\mu_1 - \mu_2 = 0$

Ha is: "The population means are not the same, i.e., $\mu_1 - \mu_2 \neq 0$
Alpha is to be set at 0.05.

$$Sp = \sqrt{\frac{(12-1)15^2 + (15-1)19^2}{12 + 15 - 2}} = 17.354$$

$$\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = 17.354 \sqrt{\frac{1}{12} + \frac{1}{15}} = 6.721$$

In the two-sample t-test, the t-statistics are retrieved by subtracting the difference between the two sample means

from the null hypothesis, which is $\mu_1 - \mu_2$ is zero.

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\overline{X}_1 - \overline{X}_2}}$$

$$= \frac{(92 - 84) - 0}{6.721} = 1.19$$

Looking up t-tables (using spreadsheet software, such as Excel's TINV function, is easiest), one finds that the critical value of t is 2.06. Again, this means that if the standardized difference between the two sample means (and that is exactly what the t value indicates) is larger than 2.06, it can be concluded that there is a significant difference between population means.

Here, 1.19 is less than 2.06; thus, it is the null hypothesis that $\mu_1 - \mu_2$ = 0.

Below is the output from statistical software using the same data:

Table 4: Two-Sample T-Test and Confidence Interval for A Sample and B Sample				
	N	Mean	Standard Deviation	SE Mean
1 Sample	12	92.0	15.0	4.3
2 Sample	15	84.0	19.0	4.9

Difference = mu (1) - mu (2)
Estimate for difference: 8.00000
95% CI for difference: (-5.84249, 21.84249)
T-test of difference = 0 (vs not =): T-value = 1.19, P-value = 0.245, DF = 25
Both use pooled StDev = 17.3540

About the Author: *Chew Jian Chieh* *is a senior consultant with Valeocon Management Consulting and supports clients in China and across Asia. He has extensive experience in implementing process and organization improvements for various industries. He specializes in Lean Six Sigma, strategy development/deployment and change management. Chew JC is a Singapore national. He can be reached at jian-chieh.chew@valeocon.com .*

Published in Hypothesis Testing

Social sharing

back to top

Published in Hypothesis Testing

Social sharing

back to top