Comparison Study by Paired Means Difference

Chanoknath Sutanapong★ & Louangrath, P.I. ★★

About the author

★Chanoknath Sutanapong is an independent researcher. She may be reached by email at: chanoknath.sutanapong@gmail.com

★★ Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. For correspondence, email: Lecturepedia@gmail.com

ABSTRACT

The paper is intended to be used as a research methodology guideline for comparison studies using T and Z tests. Comparison studies may occur in the context of quality control management. Within a production batch, one can divide the batch into two parts, treating them as sample 1 and sample 2 respectively. A comparison study of these two samples would tell us whether the production has consistent quality. If there is a consistent quality in the batch, the difference of the means of the two samples would not be significant. Throughout the discussion in this paper, the confidence interval is fixed at 95%.

Keywords: Pair comparison, paired means difference

CITATION:

Sutanapong, C. and Louangrath, P.I. (2015). "Paired Comparison Study by Paired Means Difference." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 1: pp. 12-21. (Jan. – Mar. 2015).

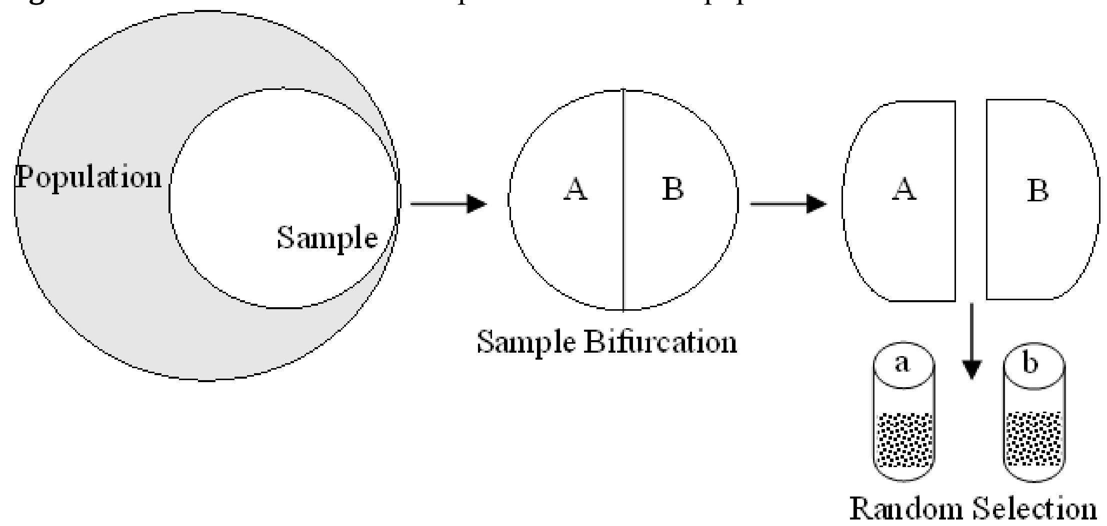
1.0 INTRODUCTION

The objective of this paper is to introduce researchers to inferences about the differences in means and variances in paired comparison design (Zimmerman, 1997). This is called the paired comparison case. If there is a lack of homogeneity among the population, a paired comparison of two experiments or samples may be studied to determine whether the heterogeneity in the population inflated the experimental error. Homogeneity is another definition for internal consistency. To achieve this objective, we introduce three tests. Two of the tests are based on T test (Mankiewicz, 2004; Fisher, 1987). One test is based on the Z test. T test is selected because it is considered the most robust (Bland, 1995) and is less likely to violate the rule of normality in large sample (Sawilowsky *et al.*, 1992). Although the Wilcoxon rank-sum test and Mann-Whitney U test are considered to have higher power than the T test (Blair and Higgins, 1980; Fay and Proschan, 2010), we select the T test because of its easier to use and the results are within acceptable bounds of scientific scrutiny. Alternatives to T test had been discussed in the literature (Sawilowsky, 2005).

The procedure for paired means difference comparison entails: (i) divide the experimental data into two groups or parts, (ii) randomly apply treatment to both. Note that applying treatment to both samples, so that they are both treatment groups. There is no control group; and (ii) apply random selection to the experiment. There are two sections in this exercise. Section 2 explains the paired comparison by using the means of two experiments. Section 3 explains the paired comparison using the variances of the two experiments. The figure below illustrates how the experiment is bifurcated into two comparable samples. In sections 2 and 3, the sample is assumed to have equal size. In Section 4, we introduce means comparison studies in a case where the means of the two samples are not equal.

We present three cases. In case 1, the two samples have equal size: $n_1 = n_2$. In this case, we suggest the use of dBar analysis. In case 2, we present a mean comparison analysis where the variances for the two samples are known and equal. In case 3, we present two samples with unknown and unequal variances. We begin with an illustration (Fig. 1) where from a population a sample is drawn (A & B).

Fig. 1. An illustration of how a sample is taken from a population



This sample is then bifurcated: separated into two segments. In each segment, a random selection is made to observe the characteristics. This research note covers two scenarios: (1) paired comparison of two *means* from the two experiments, and (2) paired comparison among two *variances* of the two experiments. The mean of the data distribution is the average of the data. The variance is the curve drawn from differences between the individual data and the mean. Compare the two means and compare the two variances of the two experiments. Note that the variance is the shape of the curve of the normal distribution of data.

The data used for illustrative purpose in this paper comes from the annual report of the IMF's economic outlook between and 2010 and 2014. The annual GDP of the 10 countries in ASEAN was used for illustrative in each case where the means are equal. For unequal means, the 10 ASEAN countries are broken down to two groups where Cambodia, Laos, Myanmar and Vietnam as one group and the remaining six countries in another group. The annual GDP for the 5 years period is presented in Table 1.

Table 1. Annual GDP of the ASEAN countries: 2010 – 2015

Country	2010	2011	2012	2013	2014
Brunei	79,302.82	82,567.71	83,659.61	82,052.74	80,221.21
Cambodia	2,459.64	2,646.49	2,840.40	3,055.27	3,280.68
Indonesia	8,432.70	8,973.56	9,554.34	10,108.43	10,661.63
Laos	4,382.25	4,761.36	5,153.34	5,576.56	6,022.04
Malaysia	20,335.84	21,498.39	22,742.22	23,630.68	25,088.81
Myanmar	3,678.78	3,932.90	4,262.68	4,655.78	5,074.26

Philippines	5,550.36	5,773.74	6,122.30	6,545.99	6,953.29
Singapore	70,657.26	75,113.21	77,690.83	81,647.56	85,227.12
Thailand	13,187.95	13,513.55	14,690.31	15,252.06	15,596.54
Vietnam	4,395.52	4,716.98	5,000.76	5,300.32	5,657.25

2.0 COMPARING PAIRED MEANS OF TWO SAMPLES WITH EQUAL SAMPLE SIZES

The first case of paired comparison studies is the *variances* comparison. The statistical model used to describe the data in the experiment described above and illustrated in Figure 1.0 is given by:

$$y_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \cdots n \end{cases} \tag{1}$$

Where y_{ij} = observations; i & j = two experiments;
μ_i = true mean for i^{th}; β_j = effect due to j^{th} specimen; and ε_{ij} = random error with mean zero and σ_i^2 variance that is: σ_1^2 = variance from experiment 1; and σ_2^2 = variance from experiment 2.

From the above stated conditions, the paired difference of the j^{th} may be calculated thus:

$$d_j = y_{1j} - y_{2j} \qquad \text{where } j = 1, 2, \cdots n \tag{2}$$

The expected value of the difference is:

$$\begin{aligned} \mu_d &= E(d_j) \qquad \text{or} \\ &= E(y_{1j} - y_{2j}) \\ &= E(y_{1j}) - E(y_{2j}) \\ &= \mu_1 + \beta_j(\mu_2 + \beta_j) \\ &= \mu_1 - \mu_2 \end{aligned}$$

Thus, the expected value of the difference among paired difference is:

$$E(d_j) = \mu_1 - \mu_2 \tag{3}$$

In order to test the *null hypothesis* that $H_0 : \mu_1 = \mu_2$ is equivalent to testing that:

$H_0 : \mu_d = 0$
$H_1 : \mu_d \neq 0$

The test statistic used for this hypothesis testing is:

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}} \tag{4}$$

where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_j$.

Statement (4a) is the sample mean of the differences, and the standard of the differences of the samples is given by:

$$S_d = \left[\frac{\sum_{i=1}^{n} (d - \bar{d})^2}{n-1} \right]^{1/2} = \sqrt{\frac{\sum_{i=1}^{n} (d - \bar{d})^2}{n-1}} \tag{5}$$

The *null hypothesis* which states that $H_0 : \mu_d = 0$ is rejected if $|t_0| > t\alpha/2, n-1$.

This type of experiment is called a *paired t-test model* because the observations from the factor levels are "paired" on each experimental unit. A data set consisting of a paired experiment to illustrate the point is in order. Assume that the following data sets comprised the paired samples:

Table 1A. ASEAN-10 dividing into 2 groups of equal sample size

Country	2010	2011	2012	2013	2014
Brunei	79,302.82	82,567.71	83,659.61	82,052.74	80,221.21
Cambodia	2,459.64	2,646.49	2,840.40	3,055.27	3,280.68
Indonesia	8,432.70	8,973.56	9,554.34	10,108.43	10,661.63
Laos	4,382.25	4,761.36	5,153.34	5,576.56	6,022.04
Malaysia	20,335.84	21,498.39	22,742.22	23,630.68	25,088.81

The 10 countries are divided into half of 5 countries each based on their alphabetical listing. The first half is listed in table 1A and the second half is listed in Table 1B.

Table 1B. ASEAN-10 dividing into 2 groups of equal sample size

Country	2010	2011	2012	2013	2014
Myanmar	3,678.78	3,932.90	4,262.68	4,655.78	5,074.26
Philippines	5,550.36	5,773.74	6,122.30	6,545.99	6,953.29
Singapore	70,657.26	75,113.21	77,690.83	81,647.56	85,227.12
Thailand	13,187.95	13,513.55	14,690.31	15,252.06	15,596.54
Vietnam	4,395.52	4,716.98	5,000.76	5,300.32	5,657.25

The difference is each year for each pair is listed in Table 1C, i.e. (Brunei – Myanmar = D1), (Cambodia – Philippines = D2), (Indonesia – Sigapore = D3), Laos – Thailand = D4) and (Malaysia – Vietnam = D5).

Table 1C. ASEAN's GDP paired difference

Country	2010	2011	2012	2013	2014
D1	75,624.04	78,634.81	79,396.93	77,396.96	75,146.95
D2	-3,090.72	-3,127.25	-3,281.90	-3,490.72	-3,672.61
D3	-62,224.56	-66,139.65	-68,136.49	-71,539.13	-74,565.49
D4	-8,805.70	-8,752.19	-9,536.97	-9,675.50	-9,574.50
D5	15,940.32	16,781.41	17,741.46	18,330.36	19,431.56
Mean	33,137.07	34,687.06	35,618.75	36,086.53	36,478.22
SD	33,323.23	35,035.42	35,424.08	35,492.07	35,484.21
T	2.22	2.21	2.25	2.27	2.30
T*	2.13	2.13	2.13	2.13	2.13

For our purpose, we will ignore the sign and use the absolute value of the difference of the monetary value. For each year, we can calculate the mean and standard deviation as shown in table 1 C.

Assume that the confidence interval of 0.95 is used, the standard random error is 0.05 for both tails. One tail of the random error is 0.025. Compute to reject the *null hypothesis* H_0 if $|t_0| > t_{0.025,4} = 2.13$. In all pairs for all years, the observed value for T is greater than 2.13. The differences among the five pairs are significant.

3.0 COMPARING PAIRED VARIANCES OF TWO EXPERIMENTS

The second case of paired comparison is the analysis of variances of two experiments. Continuing with the paired comparison experiment, we want to test the hypothesis that the difference between the variances of two experiments is a constant, that is:
$H_0 : \sigma^2 = \sigma_0^2$ and $H_1 : \sigma^2 \neq \sigma_0^2$. The test statistic is the chi-square test:

$$\chi_0^2 = \frac{SS}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \qquad (6)$$

where $SS = \sum_{i=1}^{n} \left(y_i - y \right)^2$ is the corrected sum of squares of the sample observation. The reference distribution for χ_0^2 is the chi-square distribution with $n-1$ degree of freedom. The decision rule used for hypothesis testing is:

Reject the *null hypothesis* if $\chi_0^2 > \chi_{\alpha/2,n-1}^2$
Reject the *null hypothesis* if $\chi_0^2 < \chi_{1-(\alpha/2),n-1}^2$

where $\chi_{\alpha/2,n-1}^2$ and $\chi_{1-(\alpha/2),n-1}^2$ are upper $\alpha/2$ and lower $1-(\alpha/2)$ percentage points of χ^2 distribution with $n-1$ degrees of freedom.

The $100(1-\alpha)$ percentage confidence interval on σ^2 is:

$$\frac{(n-1)S^2}{\chi_{\alpha/2,n-1}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-(\alpha/2),n-1}^2} \qquad (7)$$

If the independent random sample of size n_1 and n_2 are taken from the populations 1 and 2, the statistic for: $H_0 : \sigma^2 = \sigma_0^2$ and $H_1 : \sigma^2 \neq \sigma_0^2$ is the ratio of the sample variance which is given by:

$$F_0 = \frac{S_1^2}{S_2^2} \qquad (8)$$

The appropriate reference distribution is the F-distribution with $n-1$ degrees of freedom numerator. The decision rule used is: (i) Reject H_0 if $F_0 > F_{\alpha/2,n_1-1,n_2-1}$ a and (ii) Reject H_0 if

$F_0 < F_{1-(\alpha/2),n_1-1,n_2-1}$ where $F_{\alpha/2,n_1-1,n_2-1}$ and $F_{1-(\alpha/2),n_1-1,n_2-1}$ denote the upper $\alpha/2$ and $1-(\alpha/2)$ percentage points of the F-distribution with n_1-1 and n_2-1 degree of freedom. The upper tail and lower tails are related by:

$$F_{1-\alpha,v\ ,v_2} = \frac{1}{F_{\alpha,v_1,v_2}} \tag{9}$$

The result of the F test for variances comparison is presented in Table 2 below. Since the variance of the GDP is unreasonably large, we presented the variance in a log form: ln(var).

Table 2A. Logged Variance of the first group of ASEAN

Country	2010	2011	2012	2013	2014
Brunei	79,302.82	82,567.71	83,659.61	82,052.74	80,221.21
Cambodia	2,459.64	2,646.49	2,840.40	3,055.27	3,280.68
Indonesia	8,432.70	8,973.56	9,554.34	10,108.43	10,661.63
Laos	4,382.25	4,761.36	5,153.34	5,576.56	6,022.04
Malaysia	20,335.84	21,498.39	22,742.22	23,630.68	25,088.81
Variance 1	20.76	20.84	20.86	20.88	20.74

To reduce the level of numbers into a more manageable scale, we descale the values by taking its log. The log values are presented in Table 2B.

Table 2B. Logged Variance of the second group of ASEAN

Country	2010	2011	2012	2013	2014
Myanmar	3,678.78	3,932.90	4,262.68	4,655.78	5,074.26
Philippines	5,550.36	5,773.74	6,122.30	6,545.99	6,953.29
Singapore	70,657.26	75,113.21	77,690.83	81,647.56	85,227.12
Thailand	13,187.95	13,513.55	14,690.31	15,252.06	15,596.54
Vietnam	4,395.52	4,716.98	5,000.76	5,300.32	5,657.25
Variance 2	20.54	20.66	20.73	20.82	20.91

The F test result is presented in Table 3 below. The result of the F test is then converted back to the original scale before comparing the value to the theoretical F at degree of freedom 4. Under the F test, there is no significant difference among the pairs. The result in table 3 contradicts the findings reported in Table 1C. We attribute this difference to the fact that in the T-bar analysis reported in Table 1C, we looked at the paired difference of the annual GDP values whereas in the F test reported in Table 3, we look at the variance of the whole group for each year.

Table 3. F test result of 5 ASEAN pairs constructed from 10 ASEAN countries

Country	2010	2011	2012	2013	2014
Variance 1	20.76	20.84	20.86	20.88	20.74
Variance 2	20.54	20.66	20.76	20.82	20.91
F	1.01	1.01	1.01	1.00	0.99
Exp(F)	2.75	2.74	2.74	2.72	2.70
F*(4,4)	6.39	6.39	6.39	6.39	6.39

So far, we examined two paired comparison studies: the means comparison and the variances comparison. In the mean comparison, use the T-distribution for testing the confidence interval. In the variances comparison, use the F-distribution for testing the confidence interval.

When dealing with experimental data, there are two situations: normal distribution and non-normal distribution. In a case of normal distribution, use the Z-table as the test statistic. In a non-normal case with *on sample*, use the chi-square distribution. In a non-normal case with a paired sample, use the F-distribution. In paired sample analysis, there are two approaches: (i) means difference analysis using the T-distribution table, and (ii) variances analysis using the F-distribution table.

4.0 MEANS COMPARISON WITH VARIANCE KNOWN AND EQUAL

The second scenario introduces the mean comparison analysis in a case where there are two populations with variances known and equal. The purpose of the study is to determine whether the differences among the means of two populations are statistically different if their variances are known and equal. The usefulness of the knowledge gained from this exercise has practical application in business management, particularly in quality control.

In quality control management, the manager is dealing with uniformity of finished products. If the product comes from the same production process, it should be the same. If they are many batches of the same product being produced, all batches should be the same. The lack of uniformity of products of the same or several batches under the same production lot order leads to the conclusion that the production process lacks standard. The absence of standard in production leads to poor quality. Means comparison of two populations with known and equal variances is one means for quality control. The figure below illustrates the set up of the two populations, i.e. two production batches.

Given two populations with known means, the variances of these two populations are known and equal. The objective is to determine whether the differences in means are significant; if their means difference is significant, it means that the two populations are different.

The case involves two populations and population comparison; therefore, the test involves Z-equation. The test statistics is given by:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{10}$$

Recall the two groups of ASEAN countries: Group 1 (Cambodia, Laos, Myanmar, and Vietnam) and Group 2 (Brunei, Indonesia, Malaysia, Philippines, Singapore and Thailand). The data for the two groups are presents in Tables 4A and 4B below.

Table 4A. Group 1 of ASEAN

Country	2010	2011	2012	2013	2014
Cambodia	2,459.64	2,646.49	2,840.40	3,055.27	3,280.68
Laos	4,382.25	4,761.36	5,153.34	5,576.56	6,022.04
Myanmar	3,678.78	3,932.90	4,262.68	4,655.78	5,074.26
Vietnam	4,395.52	4,716.98	5,000.76	5,300.32	5,657.25
Mean	3,729.05	4,014.43	4,314.30	4,646.98	5,008.56
μ	2,759.80	2,962.04	3,188.84	3,444.49	3,713.25

Table 4B. Group 2 of ASEAN

Country	2010	2011	2012	2013	2014
Brunei	79,302.82	82,567.71	83,659.61	82,052.74	80,221.21
Indonesia	8,432.70	8,973.56	9,554.34	10,108.43	10,661.63
Malaysia	20,335.84	21,498.39	22,742.22	23,630.68	25,088.81
Philippines	5,550.36	5,773.74	6,122.30	6,545.99	6,953.29

Singapore	70,657.26	75,113.21	77,690.83	81,647.56	85,227.12
Thailand	13,187.95	13,513.55	14,690.31	15,252.06	15,596.54
Mean	32,911.16	34,573.36	35,743.27	36,539.58	37,291.43
μ	5,630.17	5,896.53	6,631.22	7,210.83	7,810.69

The next step is to calculate sigma for the entire group of 10 countries. The results of the relevant variables are presented in Table 5 below for the purpose of calculating T.

Table 5. Variables for calculating T for mean pair comparison under equation 10

Country	2010	2011	2012	2013	2014
Mean1	3,729.05	4,014.43	4,314.30	4,646.98	5,008.56
μ1	2,759.80	2,962.04	3,188.84	3,444.49	3,713.25
Mean2	32,911.16	34,573.36	35,743.27	36,539.58	37,291.43
μ2	5,630.17	5,896.53	6,631.22	7,210.83	7,810.69
σ pool	32,055.56	33,661.39	34,294.48	34,619.81	34,868.51
T	1.27	1.27	1.26	1.26	1.25

The result in table 5 shows that there are no significant difference between the GDP of group 1 and group 2 in the years 2010, 2011, 2012, 2013 and 2014.

5.0 MEANS COMPARISON WITH VARIANCE UNKNOWN AND UNEQUAL

The learning objective of this method is to become familiar with means comparison studies. The scenario involves the taking of two samples from two populations. From prior studies of statistics, if two populations have equal means, then they might also have equal variance; however, sometimes, they might not have equal variances. In experimental design, the researcher may have a product and would want to gauge the consumer's response to the product. In so doing, the researcher may take a sample from two populations. It is suspected that these populations are different. The only fact given is that their variances are unknown and are unequal: $\sigma_1^2 \neq \sigma_2^2$. Two samples are taken from the populations; it is assume that they are different because they have different variances: $S_1^2 \neq S_2^2$. The variances of the two samples are given by:

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2}{n_1 - 1} \quad \text{and} \quad s_2^2 = \frac{\sum_{i=1}^{n_2}(x_i - \overline{x}_2)^2}{n_2 - 1} \qquad (11)$$

The test statistic is given by:

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (12)$$

The degrees of freedom are given by:

$$v = \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{s_1^4}{n_1^2 (n_1 + 1)} + \frac{s_2^4}{n_2^2 (n_2 + 1)}} \right] - 2 \qquad (13)$$

The objective of the test is to determine whether the two samples are *indeed different* since they are drawn from two different populations. The test statistic in equation 12 differs from the test used in equation 10 in that in equation 10, the test uses the Z distribution as the reference distribution and the standard deviation was the pooled standard deviation. In equation 12, in place of the pool standard deviation, the sample variances are used. The result of equation 12 is summarized in table 6.

Table 6. Variables for calculating T for mean pair comparison under equation 12

Country	2010	2011	2012	2013	2014
Mean1	3,729.05	4,014.43	4,314.30	4,646.98	5,008.56
μ1	2,759.80	2,962.04	3,188.84	3,444.49	3,713.25
Mean2	32,911.16	34,573.36	35,743.27	36,539.58	37,291.43
μ2	5,630.17	5,896.53	6,631.22	7,210.83	7,810.69
σ pool	32,055.56	33,661.39	34,294.48	34,619.81	34,868.51
T	1.23	1.23	1.23	1.23	1.22

The results under equation 12 are comparable to those of equation 10. The threshold for the 95% confidence interval is 1.64. Since none of the observed value of T under equation 10 and 12 exceeds 1.64, we conclude that the difference of the means between two groups is statistically insignificant. Not we are using 1.64 as the threshold value for T because ideally the population and the sample are equivalent at T = 1.64 and Z = 1.65 (Box *et al.*, 1987).

6.0 CONCLUSION

This paper is intended to provide a tool for researchers to do paired means comparison. This instructional guide has practical value for researcher. In industry practice, for instance in quality control, it is common for practitioner to face paired mean comparison. With a known sample, one can divide the sample into two batches and employed the paired means comparison under equations 4, 10 and 12 to determine whether they are significantly different. Significance difference means inconsistency in quality or that the sample came from a mixed source or that the machine produces inconsistent output; it could even tell us that there is a system failure. These three methods are useful tools for industrial production management.

REFERENCES

Bland, Martin (1995). *An Introduction to Medical Statistics*. Oxford University Press. p. 168. ISBN 978-0-19-262428-4.

Box, George; Hunter, William; Hunter, J. Stuart (1987). *Statistics for Experimenters*. pp. 66–67. ISBN 978-0471093152.

Blair, R. Clifford; Higgins, James J. (1980). “A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to That of Student's t Statistic Under Various Nonnormal Distributions.” *Journal of Educational Statistics*. **5**(4): 309–335. doi:10.2307/1164905. JSTOR 1164905.

Fay, Michael P.; Proschan, Michael A. (2010). "Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules." *Statistics Surveys*. **4**: 1–39. doi:10.1214/09-SS051. PMC 2857732 Freely accessible. PMID 20414472.

Mankiewicz, Richard (2004). The Story of Mathematics (Paperback ed.). Princeton, NJ: Princeton University Press. p. 158. ISBN 9780691120461.

Fisher Box, Joan (1987). "Guinness, Gosset, Fisher, and Small Samples". Statistical Science. 2 (1): 45–52. doi:10.1214/ss/1177013437. JSTOR 2245613.

Sawilowsky, Shlomo S. (2005). "Misconceptions Leading to Choosing the t Test Over The Wilcoxon Mann–Whitney Test for Shift in Location Parameter." *Journal of Modern Applied Statistical Methods*. **4**(2): 598–600.

Sawilowsky, Shlomo S.; Blair, R. Clifford (1992). "A More Realistic Look at the Robustness and Type II Error Properties of the t Test to Departures From Population Normality." *Psychological Bulletin*. 111 (2): 352–360.

doi:10.1037/0033-2909.111.2.352.

Zimmerman, Donald W. (1997). "A Note on Interpretation of the Paired-Samples t Test." Journal of Educational and Behavioral Statistics. 22 (3): 349–360. doi:10.3102/10769986022003349. JSTOR 1165289.