

## The Paired T-Test Using MINITAB

*By Keith M. Bower, M.S.*

The paired t-test procedure is used to compare the mean difference between two populations when one believes that some dependency exists. For example, if one were to test pulse rates for a group of individuals prior to, then upon completion of, a fitness regime, it would be appropriate to compare the pulse rates for each individual. What follows is an example of the paired (or dependent) t-test procedure using the popular statistical software package, MINITAB<sup>®</sup>. We may wish to test if the mean difference is significantly different from zero, i.e. test  $H_0 : \mu_{\text{difference}} = 0$ , versus an alternative hypothesis such as  $H_1 : \mu_{\text{difference}} \neq 0$ .

The t-test was originally developed by the English statistician William Sealy Gosset (1876-1937), whilst working for a brewing company in Ireland. Gosset published under the pseudonym “Student” [1908]<sup>1</sup> which is why it is frequently referred to as “Student’s” t-test.

For this example we shall consider a set of data from Stichler, Richey and Mandel [1952]<sup>2</sup> that deal with measurements of tire wear (thous mi). Each tire was subject to measurement by two different methods, the first based on weight loss, and the second based on groove wear.

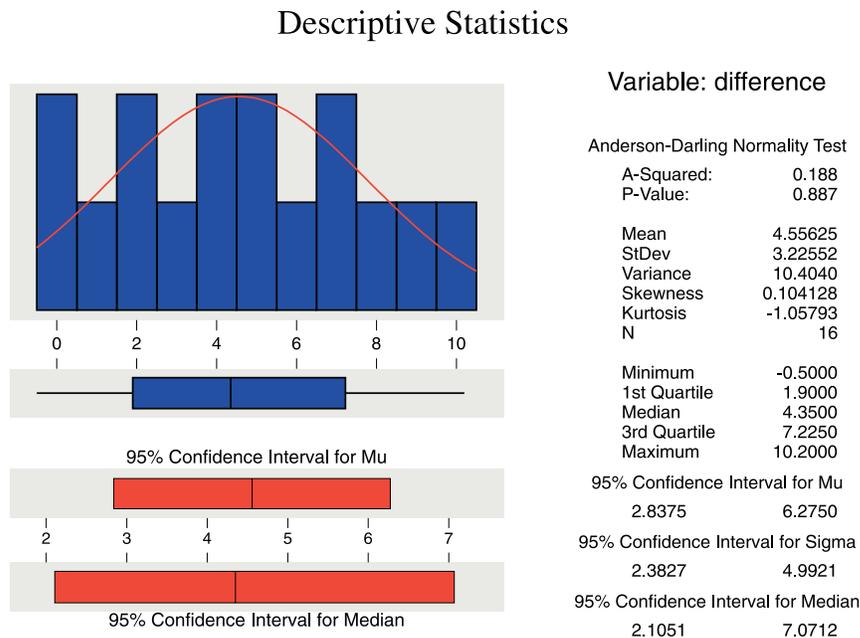
An assumption for the paired t-test procedure is that the distribution in which the differences analyzed come from is Normal (a.k.a. Gaussian). We may therefore wish to create a column for the differences between the two measurement systems, and investigate the distributional properties. The Anderson-Darling (A-D) Normality Test illustrated in Figure 1 shows that we are unable to reject the null hypothesis,  $H_0$ : data follow a Normal distribution vs.  $H_1$ : data do not follow a Normal distribution, at the  $\alpha = 0.10$  significance level. This is because the p-value for the A-D test is 0.887, which is greater than 0.10, a frequently used level of significance for such a hypothesis test (as opposed to the more traditional 0.05 level). Note that the p-value is the probability of incorrectly rejecting the null hypothesis (i.e. making a Type I error).

---

<sup>1</sup> “Student” (1908). “The Probable Error of a Mean”. *Biometrika*, 6.

<sup>2</sup> Stichler, R.D., Richey, G.G., Mandel, J. (May 1953). “Measurement of Treadwear of Commercial Tires”, *Rubber Age*, 73:2.

**Figure 1**



As is discussed by Hogg and Ledolter [1987]<sup>3</sup>, when the assumption of Normality is violated, the t-test procedure still works well if the underlying distribution (in this case for the differences) is symmetric, unimodal and continuous. If the values (in this case the differences) are seriously skewed, it may appropriate to use a nonparametric procedure, such as the one-sample sign test. However, provided that the precaution of randomization is employed, use of the paired t-test procedure without concern as to the form of the underlying distribution is a valid approach. For more information, the interested reader is referred to Box, Hunter and Hunter [1978]<sup>4</sup>.

We may now perform a paired t-test. As the output in Figure 2 exhibits, we are able to reject the null hypothesis,  $\mu_{\text{difference}} = 0$  at the  $\alpha = 0.05$  level of significance as the p-value is less than 0.05. In fact, the evidence strongly suggests that there is a difference between the two methods (the p-value is very low). From the data in this study we can conclude that, on average, the weight loss method gives higher measurements than the groove wear method.

<sup>3</sup> Hogg, R.V., Ledolter, J. (1987). Applied Statistics for Engineers and Physical Scientists. Macmillan Publishing Company, NY.

<sup>4</sup> Box, G.E.P., Hunter, W.G., Hunter, J.S.. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. John Wiley and Sons, Inc.

**Figure 2**

**Paired T-Test and CI: WGT, GRO**

Paired T for WGT - GRO

	N	Mean	StDev	SE Mean
WGT	16	28.16	9.53	2.38
GRO	16	23.60	7.94	1.99
Difference	16	4.556	3.226	0.806

95% CI for mean difference: (2.837, 6.275)

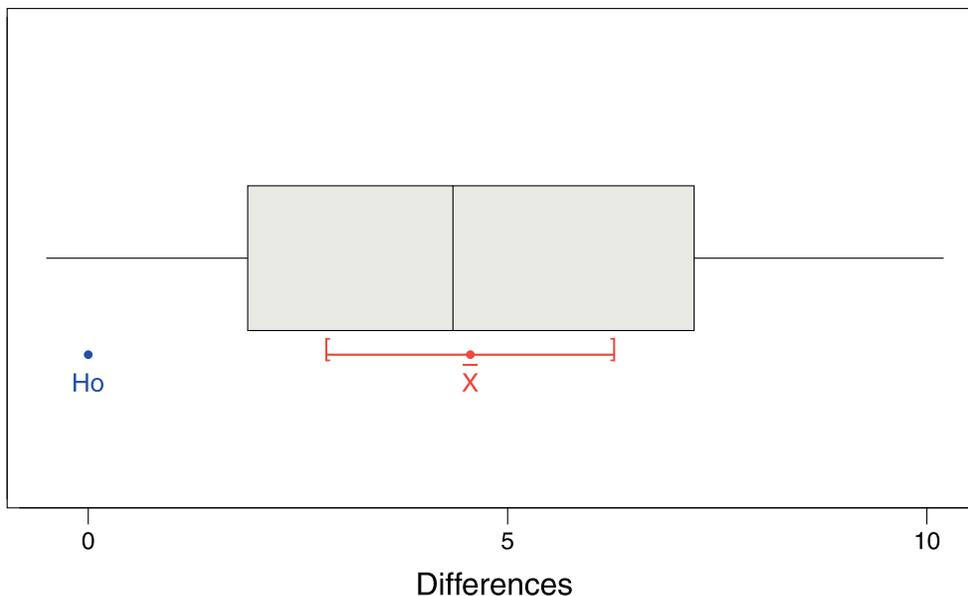
T-Test of mean difference = 0 (vs not = 0): T-Value = 5.65 P-Value = 0.000

It would be useful for us to have some appreciation as to a reasonable range of values for what the true mean difference may be for the two types of measurement systems. Here we tested the null hypothesis using a significance level of  $\alpha = 0.05$ . A  $100*(1 - \alpha)\%$  confidence interval, i.e. a 95% confidence interval, is also reported in the MINITAB output in Figure 2 and displayed in Figure 3. We may be 95% confident that the true mean difference in tire tread measurements is between 2.837 and 6.275. One should always keep in mind that though a difference may be statistically significant (as in this case), there may not be a practical difference.

**Figure 3**

**Boxplot of Differences**

(with Ho and 95% t-confidence interval for the mean)



In conclusion, one finds that the paired t-test can be used to investigate the mean difference between two dependent populations. If the means from two independent populations are to be investigated, the two-sample t-test may be used. For the comparison of means from two or more independent populations, the ANOVA procedure<sup>5</sup> may be used.

*Keith M. Bower, M.S. is a Technical Training Specialist with Minitab, Inc. His main interests are in continuous quality improvement techniques, especially control charting, as well as business and healthcare statistical methodologies. Keith is a member of the American Society for Quality and the American Statistical Association.*

---

<sup>5</sup> Bower, K.M. (February 2000). "The ANOVA Procedure using MINITAB". Scientific Computing and Instrumentation.