

2017

A Review of the Multiple-Sample Tests for the Continuous-Data Type

Dewi Rahardja

U.S. Department of Defense, rahardja@gmail.com

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Rahardja, D. (2017). A review of the multiple-sample tests for the continuous-data type. *Journal of Modern Applied Statistical Methods*, 16(1), 127-136. doi: 10.22237/jmasm/1493597220

A Review of the Multiple-Sample Tests for the Continuous-Data Type

Cover Page Footnote

Disclaimer Statement: This research represents the author's own work and opinion. It does not reflect any policy nor represent the official position of the U.S. Department of Defense nor any other U.S. Federal agency.

A Review of the Multiple-Sample Tests for the Continuous-Data Type

Dewi Rahardja

United States Department of Defense
Fort Meade, MD

For continuous data, various statistical hypotheses testing methods have been extensively discussed in the literature. In this article a review is provided of the multiple-sample continuous-data testing methods. It includes traditional methods, such as the two-sample t -test, Welch ANOVA test, etc., as well as newly-developed ones, such as the various Multiple Comparison Procedure (MCP). A roadmap is provided in a figure or diagram format as to which methods are available in the literature. Additionally, the implementation of these methods in popular statistical software packages such as SAS is also presented. This review will be helpful to determine which continuous-data testing method (along with the corresponding SAS code) are available to use in various fields of study, both for the design phase of a study in prospective study, cross-sectional, or retrospective study analysis and the analysis phase.

Keywords: Two-sample t -test, one-way ANOVA, Satterthwaite, degrees of freedom, Welch ANOVA, Wilcoxon rank-sum test, Kruskal-Wallis test, paired t -test, multiple comparison procedure (MCP)

Introduction

In many real-world applications, such as data in clinical trials, financial data, epidemiology, sociology, etc., we often encounter data with outcome (or response) variables that are continuous in nature. If a random variable can take any value within an interval or continuum, it is called a continuous random variable. For example, diastolic blood pressure, amount of dollar expenses, height, weight, cholesterol level, air pollutant level, etc. are usually considered continuous random variables because they can take any value within certain intervals, even though the observed measurement is limited by the accuracy of the measuring device. Due to the nice asymptotic math/stat properties, the Normal distribution is the most commonly-used continuous distribution in the fields of clinical research, finance, epidemiology, sociology, along with many others.

Dewi Rahardja is a Statistician. Email her at: rahardja@gmail.com.

A REVIEW OF CONTINUOUS-DATA TESTING METHODS

Without loss of generality (WLOG), the standard (classical or frequentist) large sample (asymptotic) theory is derived using the underlying assumptions of independent, normally-distributed random variables with homogeneous (i.e., equal) variance.

A frequent task in data analysis is to check these three assumptions (in the order of: independent, normal, equal variance) for the outcome measure or response variable, and then to determine what test is suitable/appropriate for a dataset.

Such continuous-data outcome measure or response variables (or dependent variables) can occur both in randomized controlled trials and in observational studies. The predictor or covariate (or independent variable) is the terminology used for both continuous and categorical variable. However specifically, the predictor is called a grouping variable (or factor) for a discrete/categorical predictor. Typically, this grouping variable can have one, two, or multiple levels. The common (or generic) statistical terms used are one-, two-, and multiple-sample testing methods for one, two, and multiple levels of this one factor (or grouping variable).

To date, there is no literature that comprehensively presents and summarizes the review of the various one-sample, two-sample, and multiple-sample tests for the continuous-data type of response variable (or outcome measure) with one grouping variable (factor) of multiple levels. Hence in our line of (statistical) practice, we often find both statistician and non-statistician practitioners, investigators, and researchers get confused/mixed-up about the method, model, and hypothesis to use. To close this confusion gap, this article will be a very useful basic guidance/roadmap to both statisticians and non-statisticians in various fields of study.

For the categorical-data type (of outcome measure or response variable), Rahardja, Yang, and Zhang (2016) have provided a comprehensive review, also in a roadmap format, along with the corresponding translation/implementation of those methods in popular and professional statistical software packages, such as SAS and/or R.

Hypothesis Testing

First, we begin with the popular one-sample mean test (for a normal population): the one-sample z -test and the one-sample t -test (not listed on Table 1 nor Figure 1). WLOG, consider the simplest case: a continuous response variable (or outcome measure), Y , with one grouping variable (or factor), X , as the discrete

covariate or factor (or predictor). This single factor has only one level (i.e., $X = 1$). For this very basic/simplest model, the objective is to model the expected value of a continuous random variable, Y , as a linear function of the discrete predictor or factor, X , and hence $E(Y_i) = \mu_X$. This basic/simplest model has only one factor with one level (i.e., $X = 1$); therefore $E(Y_i) = \mu$. Hence, this (generic) model structure can be written as $Y_i = \mu_X + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$, for $i = 1, 2, \dots, n$ observations (which is a statistical linear model which is linear in the parameter, μ). Essentially, this model structure can be simplified as the mean model (for one factor), $Y_i = \mu + \epsilon_i$ where $i = 1, 2, \dots, n$ observations. For this (generic) basic model the assumptions are that Y is normally distributed, errors are normally distributed and independent with constant/homogeneous variance σ^2 , i.e. $\epsilon_i \sim N(0, \sigma^2)$, while X is fixed (i.e., $X = 1$); see Casella (2008).

Theoretically, with a known standard deviation (σ), the standard one-sample z -test can be used to test the null hypothesis, $H_0: \mu = 0$, versus the alternative hypothesis, $H_1: \mu \neq 0$. However, practically, the standard deviation (σ) is unknown, and hence the one-sample t -test can be used to test the same aforementioned hypothesis.

Second, consider the two-sample (and subsequently, multiple-sample) mean test (see Figure 1), depending on the assumptions of the response variable (or outcome measure). Consider the case: a continuous response variable (or outcome measure), Y , with one grouping variable (or factor), X , as the discrete covariate or factor (or predictor). This single factor has two (or more) levels (e.g., $X = 0$ for the placebo group, or for $X = 1$ the drug A group, or $X = 2$ for the drug B group, etc.), and can be written as an indicator function/variable. This model structure can be written as the so-called cell means model (for one factor), $Y_{ij} = \mu_i + \epsilon_{ij}$, where $i = 1, 2, \dots, k$ groups (i.e., the i^{th} level of that one factor), and $j = 1, 2, \dots, n$ observations; see Casella (2008). The model assumptions are that Y_{ij} is normally distributed, errors are normally distributed and independent with constant/homogeneous variance σ^2 , i.e. $\epsilon_{ij} \sim N(0, \sigma^2)$; X is a fixed indicator function/variable (i.e., $X = 0, 1$, etc.); and μ_i is the unknown theoretical/population mean for all of the observations at level i .

The generic hypothesis testing for two means can be written as $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$, and for multiple means it can be generalized as $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ versus H_1 : at least one mean is different than the rest.

Next, consider how to implement these methods (in Figure 1) in popular statistical software packages, such as SAS (see Table 1). The SAS PROC TTEST, or the TTEST procedure, performs t -tests for one-sample, two-sample, and paired observations (see Table 1 and Figure 1). The one-sample t -test compares the mean

A REVIEW OF CONTINUOUS-DATA TESTING METHODS

of the sample to a given number (which you supply, and typically is zero). The dependent-sample or paired t -test compares the difference in the means from the two variables to a given number (usually 0) while taking into account the fact that the scores are not independent (i.e., paired scores or data); see David and Gunnink (1997). The independent samples t -test (or two-sample t -test) compares the difference in the means from the two groups to a given value (usually 0). In other words, it tests whether the difference in the means is 0.

When there are multiple levels within that one factor (or one way) model (of the cell means model), alternatively the model can be written as the effect model to test the effect of the multiple levels (i.e., multiple-sample test); similarly for the two levels (i.e., two-sample test). The effect model is used to separate the baseline mean effect from the groups' or levels' effect: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $i = 1, 2, \dots, k$ groups (i.e., the i^{th} level of that one factor), and $j = 1, 2, \dots, n$ observations; and to test the multiple-level effect, $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k$. The SAS procedure PROC ANOVA can be used for such multiple-sample test.

When the response variable (or outcome measure) holds the assumptions of independent, normally distributed with homogeneous (equal variance), then the One-Way ANOVA method can be implemented via the SAS procedure, PROC ANOVA with means statement, using the option /hovtest. See Zimmerman (2004), who discussed preliminary tests of equality of variances.

Similarly, when the response variable (or outcome measure) holds the assumptions of independent and normally distributed with non-homogeneous (or heterogeneous or unequal) variances, then the Welch (1947) ANOVA method can be implemented via the SAS procedure, PROC ANOVA with means statement, using the option /welch.

Wilcoxon (1945) and Mann and Whitney (1947) proposed a distribution-free model (i.e., nonparametric statistical methods) where the null hypothesis can be written as $H_0: F_1(X) = F_2(X)$ where $F_i(X)$ is the distribution function for sample $i = 1, 2$. This null hypothesis is to test whether the two population distributions are identical by using the sum of the ranks in sample 1 and sample 2. The test statistic is called the Wilcoxon rank-sum test (Mann-Whitney test). Alternatively, Zhao, Rahardja, and Qu (2008) considered quantifying the difference between the two groups, and defined the hypothesis in terms of the competing probability, $\pi = \Pr(X > Y) + 0.5 \Pr(X = Y)$, where X and Y are random variables with cumulative distribution functions (CDFs) F_X and F_Y , respectively. Then the following null hypothesis indicates there is no difference between the two groups: $H_0: \pi = 0.5$. Here the SAS procedure used is the PROC NPAR1WAY with Wilcoxon statement. For the distribution-free model (i.e., nonparametric statistical

methods) with multiple levels (multiple samples) within that one factor (or the grouping variable), the Kruskal-Wallis test of $H_0: F_1(X) = F_2(X) = \dots = F_k(X)$ can be used (Kruskal & Wallis, 1952). Here the SAS PROC NPAR1WAY can be used.

Cao and Zhang (2014) reviewed various multiple comparison procedures (MCPs). Typically these MCPs are a part of an omnibus test (a series of sequential tests). For example, if using PROC GLM yields a statistically significant result for a main effect (or for an interaction, in the case of a two-factor or more scenarios), then one could use PROC MULTTEST to conduct the (pairwise) multiple comparisons. This PROC MULTTEST gives the raw p -values adjusted by Holm, Hochberg, or false discovery rate (FDR) methods. Note that under the LSMEANS statement of the PROC GLM, the “Adjust = BON;” option indicates the Bonferroni method. Among many of the above MCPs, the most commonly-used ones are Tukey’s pairwise comparison, Bonferroni’s method, Duncan, etc., depending on the specific needs, assumptions, or objective of the practitioners/researchers. For example, Tukey’s method controls the Type I experiment-wise error rate and Bonferroni, Tukey’s Least Significant Difference (LSD), and Duncan control the Type I comparison-wise error rate. Bonferroni has a very conservative (very wide) interval, i.e., is very slow to reject the null hypothesis. Table 1 summarizes the above discussion.

Roadmap

Provided in Figure 1 is the (two-sample and multiple-sample) roadmap for practitioners and researchers to choose a suitable testing method for their continuous (outcome measure or response variable) data analysis. In Figure 1, the roadmap method is provided by whether or not the response variable (outcome measure) is independent, then by whether or not the outcome is normally-distributed data, and then, finally, by whether or not the outcome variable has homogeneous variance. Then either yes/no response variable (in each of the 3 aforementioned questions) will lead to whether the grouping variable (or factor) is two-sample for a two-level factor or is multiple-sample or k -sample (where k is greater than 2) for a multiple-level factor. Next, the corresponding SAS procedures to the suitable statistical method directed from Figure 1 can be found in the Table 1 prescription.

Conclusion

Continuous data response or outcome is very common in real-data applications such as clinical trials, financial data, epidemiology, sociology, etc. The analysis of such continuous outcome measure (or response variable) has a long history, beginning with the one-sample t -test, two-sample t -test, up to the MCP. A review of the hypothesis testing procedures that are available for various types of continuous data outcome measure (or response variable) with one grouping variable (factor) of multiple levels are reviewed, along with the corresponding statistical computing translations/implementation in SAS, the most commonly used professional statistical software for data analysis.

Disclaimer

This research represents the author's own work and opinion. It does not reflect any policy nor represent the official position of the U.S. Department of Defense nor any other U.S. Federal agency.

References

- Cao, J., & Zhang, S. (2014). Multiple comparison procedures. *The Journal of the American Medical Association*, *312*(5), 543-544. doi: 10.1001/jama.2014.9440
- Casella, G. (2008). *Statistical design*. New York, NY: Springer.
- David, H. A., & Gunnink, J. L. (1997). The paired t test under artificial pairing. *The American Statistician* *51*(1), 9-12. doi: 10.1080/00031305.1997.10473578
- Kruskal, W., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* *47*(260), 583-621. doi: 10.2307/2280779
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50-60. doi: 10.1214/aoms/1177730491
- Rahardja, D., Yang, Y., & Zhang, Z. (2016). A comprehensive review of the two-sample independent or paired binary data – With or without stratum effects. *Journal of Modern Applied Statistical Methods*, *15*(2), 215-223. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol15/iss2/16/>

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28-35. doi: 10.2307/2332510

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80-83. doi: 10.2307/3001968

Zhao, Y. D., Rahardja, D., & Qu, Y. (2008). Sample size calculation for the Wilcoxon-Mann-Whitney test adjusting for ties. *Statistics in Medicine*, 27(3), 462-468. doi: 10.1002/sim.2912

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1), 173-181. doi: 10.1348/000711004849222

Appendix A: Tables and Figures

Table 1. Listing of response variable (outcome measure) type with the appropriate hypothesis testing, test statistic, and SAS command

Response (Outcome) Type/Assumptions	Null Hypothesis (H ₀)	Test statistics	SAS command or other option
Independent, normal, homogeneous variance	$Y_{ij} = \mu_i + \epsilon_{ij}$ (cell means model) $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ (effect model) where $i = 1, 2, \dots, k$ group, $j = 1, 2, \dots, n_i$ observation		
Grouping variable: two-sample	H ₀ : $\mu_1 = \mu_2$ (cell means model) H ₀ : $\alpha_1 = \alpha_2$ (effect model)	Two-sample <i>t</i> -test (S-pooled)	PROC TTEST with class statement
Grouping variable: <i>k</i> -sample	H ₀ : $\mu_1 = \mu_2 = \dots = \mu_k$ (cell means model) H ₀ : $\alpha_1 = \alpha_2 = \dots = \alpha_k$ (effect model)	One-Way ANOVA	PROC ANOVA with means statement, using /hovtest option
Independent, normal, non-homogeneous variance	$Y_{ij} = \mu_i + \epsilon_{ij}$ (cell means model) $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ (effect model) where $i = 1, 2, \dots, k$ group, $j = 1, 2, \dots, n_i$ observation		
Grouping variable: two-sample	H ₀ : $\mu_1 = \mu_2$ (cell means model) H ₀ : $\alpha_1 = \alpha_2$ (effect model)	2-sample <i>t</i> -test (Satterthwaite exact d.f.)	PROC TTEST using /cochran option
Grouping variable: <i>k</i> -sample	H ₀ : $\mu_1 = \mu_2 = \dots = \mu_k$ (cell means model) H ₀ : $\alpha_1 = \alpha_2 = \dots = \alpha_k$ (effect model)	Welch ANOVA	PROC ANOVA using /welch option, under the means statement

Table 1, continued

Response (Outcome) Type/Assumptions	Null Hypothesis (H ₀)	Test statistics	SAS command or other option
Independent, non-normal Grouping variable: 2-sample	Distribution shapes are the same but unspecified (distribution-free model) 2 Identical Distributions: H ₀ : $F_1(X) = F_2(X)$ Difference between 2 groups using competing probability: H ₀ : $\pi = 0.5$, where $\pi = P(X_1 > X_2) + 0.5 P(X_1 = X_2)$ with random variables X_1, X_2 with CDFs F_1, F_2 , respectively	Wilcoxon rank-sum test (Mann-Whitney test)	PROC NPAR1WAY with wilcoxon statement
Grouping variable: k -sample	H ₀ : $F_1(X) = F_2(X) = \dots = F_k(X)$	Kruskal-Wallis Test	PROC NPAR1WAY
<hr/>			
Not independent			
Grouping variable: two-sample	H ₀ : $\delta = 0$ $\delta = (\mu_1 - \mu_2)$	Paired t -test	PROC TTEST with paired statement
Grouping variable: k -sample	H ₀ : $\delta_1 = \delta_2 = \dots = \delta_k$ where $\delta_i = (\mu_{i,1} - \mu_{i,2})$ $i = 1, \dots, k$	Various MCPs such as Bonferroni, Tukey's LSD, Duncan, etc. See Cao and Zhang (2014)	Omnibus Test: PROC GLM using /Adjust=BON; option, under the LSMEANS statement PROC MULTTEST

A REVIEW OF CONTINUOUS-DATA TESTING METHODS

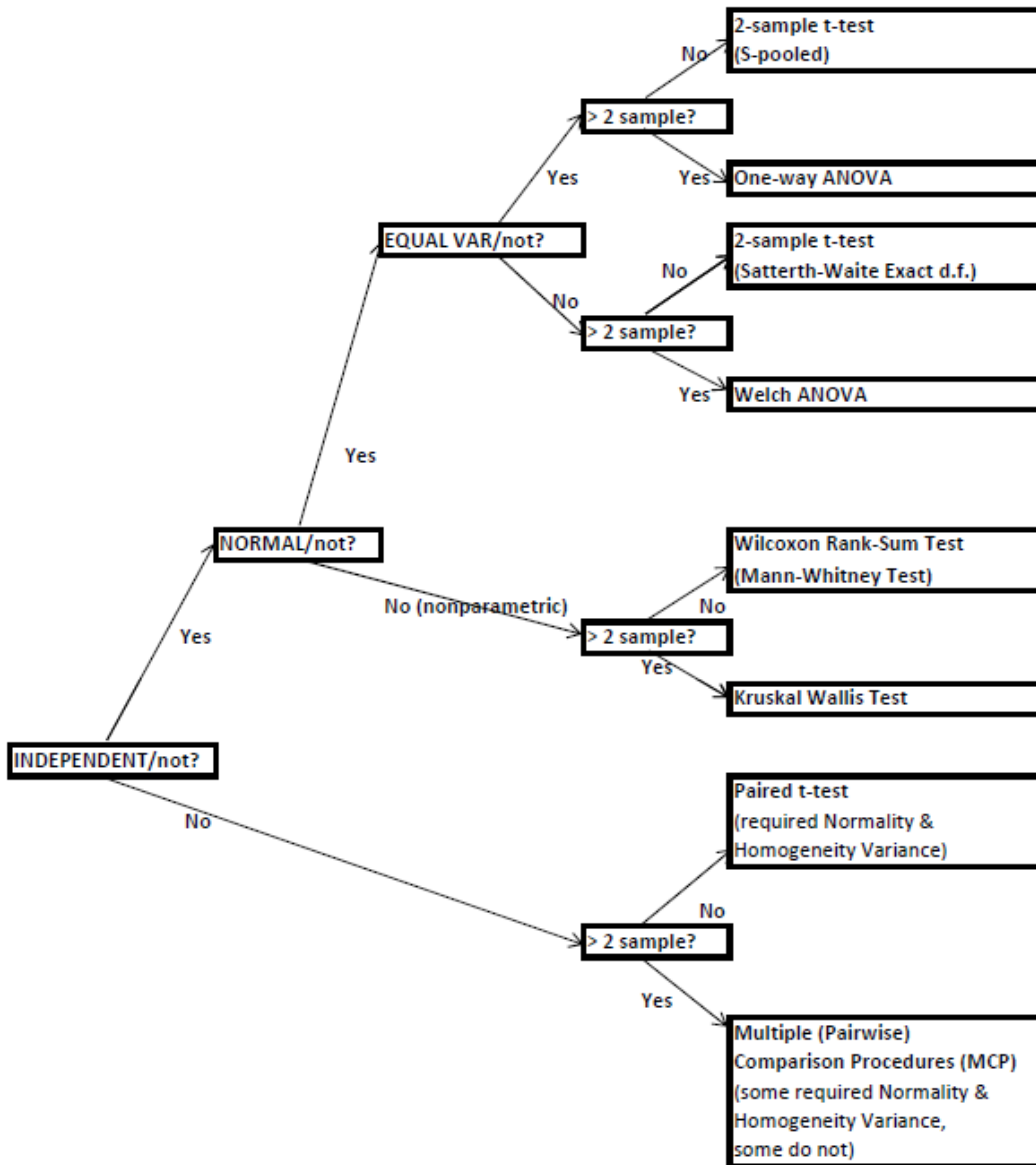


Figure 1. Continuous-data roadmap for two-sample and multiple-sample testing