AN OVERVIEW OF REMEDIAL TOOLS FOR THE VIOLATION OF PARAMETRIC TEST ASSUMPTIONS IN THE SAS SYSTEM

Chong Ho Yu, Ph.D., Arizona State University, Tempe AZ

ABSTRACT

Parametric tests are widely applied by researchers in every discipline. However, sometimes the appropriateness of their applications is in question. In a parametric test an underlying distribution is assumed, and a sample statistic is obtained to estimate the population parameter. Because this estimation process involves a sample, a sampling distribution, and a population, certain assumptions are required to ensure that all components are compatible with each other. As a matter of fact, data collected in social sciences usually violate parametric assumptions to some degree. The objectives of this paper are to briefly explain why these violations are detrimental to research and to introduce what alternatives could be used in the SAS system to rectify the situation.

INTRODUCTION

Parametric tests are widely applied by researchers in every discipline. However, sometimes the appropriateness of the application is in question. After reviewing over 400 large data sets, Micceri (1989) found that the great majority of data collected in behavioral sciences do not follow univariate normal distributions. Breckler (1990) reviewed 72 articles in personality and social psychology journals and found that only 19% acknowledged the assumption of multivariate normality, and less than 10% considered whether this assumption had been violated. Having reviewed articles in 17 journals, Keselman et al. (1998) found that researchers rarely verify that validity assumptions are satisfied and they typically use analyses that are nonrobust to assumption violations. As a matter of fact, data collected in social sciences usually violate parametric assumptions to some degree. Cliff (1996) observed that even if data collection is carefully thought out, data are often not based upon a highly parametric model. The objectives of this paper are to briefly explain why these violations are detrimental to research and to introduce what alternatives could be used in the SAS system to rectify the situation.

WHY ARE PARAMETRIC ASSUMPTIONS IMPORTANT?

In a parametric test an underlying distribution is assumed, and a sample statistic is obtained to estimate the population parameter. Because this estimation process involves a sample, a sampling distribution, and a population, certain assumptions are required to ensure that all components are compatible with each other. Take ANOVA as an example. ANOVA is a procedure of comparing means in terms of variance with reference to a normal distribution. The inventor of ANOVA, Sir R. A. Fisher (1935), clearly explained the relationship among the mean, the variance, and the normal distribution: "The normal distribution has only two characteristics, its mean and its variance. The mean determines the bias of our estimate, and the variance determines its precision" (p.42). It is generally agreed that the estimation becomes more precise as the variance becomes smaller and smaller.

To put it another way, the purpose of ANOVA is to extract precise information out of a bias, or to filter a signal out of noise. When the data are skewed (non-normal), the means can no longer reflect the central location and thus the signal is biased. When the variances are unequal, not every group has the same level of noise, and thus the comparison is invalid. More importantly, the purpose of a parametric test is to make inferences from the sample statistic to the population parameter through sampling distributions. When the assumptions are not met in the sample data, the statistic may not be a good estimation to the parameter. It is incorrect to say that the population is assumed to be normal and therefore the researcher demands the same properties in the sample. Actually, the target population is infinite in size and unknown in distribution. It may or may not possess those attributes. The required assumptions are imposed on the data because those attributes are found in sampling distributions. However, very often the acquired data do not meet these assumptions. Nevertheless, in the SAS system there are at least five sets of tools to address these problems:

1. Monte Carlo simulations (e.g., SAS/IML)
2. Non-parametric tests (e.g., PROC NONPARA1WAY)
3. Robust procedures (e.g., PROC UNIVARIATE, SAS/INSIGHT, PROC LOESS)
4. Data transformation (e.g., SAS/INSIGHT)
5. Resampling procedures (e.g., PROC MULTTEST)

Each of these tools will be discussed in the following.

MONTE CARLO SIMULATIONS: TEST OF THE TEST

If you are familiar with Monte Carlo simulations (research with dummy data), you can defend your case by citing Glass et al.'s (1972) finding that many parametric tests are not seriously affected by violation of assumptions. Indeed, it is generally agreed that the t-test is robust against mild violations of assumptions in many situations and ANOVA is also robust if the sample size is large. For this reason, Box (1953) mocked the idea of testing the variances prior to applying an F-test: "To make a preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" (p.333)

In spite of this assurance, there are still some important questions: How large should the sample size be to make ANOVA robust? How mild a violation is acceptable? How extreme is extreme? Questions like these have been extensively studied by Monte Carlo simulations. One can use the interactive matrix language in SAS/IML to study the behavior of tests (e.g., Thompson et al., 2002). Usually in a simulation, three parameters need to define, namely, the sample size, the number of replications, and the seed for random numbers. However, Monte Carlo simulations are extremely computing-intensive. Also, researchers who are not in this field will be hard pressed to "test the test" before employing any test. Nevertheless, researchers could consult the results of Monte Carlo studies to determine whether a specific parametric test is suitable to their specific data structure.

NON-PARAMETRIC TESTS

Some researchers apply non-parametric tests when assumptions are violated. As the name implies, non-parametric tests do not require parametric assumptions because interval data are converted to rank-ordered data. Examples of non-parametric tests are:

- Wilcoxon signed rank test
- Whitney-Mann-Wilcoxon (WMW) test
- Kruskal-Wallis (KW) test
- Friedman's test

In SAS/STAT, PROC NONPARA1WAY is available for comparing groups in terms of rank and scale differences. When there are two groups in the data, the tests are based on simple linear rank statistics. When the data have more than two groups, the tests are based on one-way ANOVA statistics. Both asymptotic and exact p-values are reported. This procedure also computes statistics of the empirical distribution function. This statistics indicates whether the distribution of a variable is the same across different groups. However, only asymptotic p-values are available for these tests.

Handling of rank-ordered data is considered a merit of non-parametric tests. Gibbons (1993) observed that ordinal scale data are very common in social science research and almost all attitude surveys use a 5-point or 7-point Likert scale. But the nature of this type of data is ordinal rather than interval. In Gibbons' view, non-parametric tests are considered more appropriate than classical parametric procedures for Likert-scaled data. [1] However, non-parametric procedures are criticized for the following reasons:

1. Losing precision: Edgington (1995) asserted that when more precise measurements are available, it is unwise to degrade the precision by transforming the measurements into ranked data. [2]

2. Low power: Generally speaking, the statistical power of non-parametric tests is lower than that of their parametric counterparts except on a few occasions (Hodges & Lehmann, 1956; Tanizaki, 1997).

3. Inaccuracy in multiple violations: Non-parametric tests tend to produce biased results when multiple assumptions are violated (Glass, 1996; Zimmerman, 1998).

4. Testing distributions only: Further, non-parametric tests are criticized for being incapable of answering a focused question. For example, the WMW procedure tests whether the two distributions are different in some way but does not show how they differ in mean, variance, or shape. Based on this limitation, Johnson (1995) preferred robust procedures and data transformation to non-parametric tests. Robust procedures and data transformation will be introduced next.

Taking all of the above shortcomings into account, non-parametric tests are generally not recommended

ROBUST PROCEDURES

The term "robustness" can be interpreted literally. If a person is robust (strong), he will be immune from hazardous conditions such as extremely cold or extremely hot weather, viruses, etc. If a test is robust, the validity of the test result will not be affected by poorly structured data. In other words, it is resistant against violations of parametric assumptions. Robustness has a more technical definition: If the actual Type I error rate of a test is close to the proclaimed Type I error rate, say 0.05, the test is considered robust.

Several conventional tests have some degree of robustness. For example, Satterthwaite's (1946) t-test used by SAS could compensate unequal variances between two groups. In SAS when you run a t-test, SAS can also test the hypothesis of equal variances. When this hypothesis is rejected, you can choose the t-test adjusted for unequal variances (see Table 1).

Table 1. T-test result from SAS

Variances	**T**	**DF**	**Prob>|T|**
Unequal	-0.0710	14.5	0.9444
Equal	-0.0750	24.0	0.9408

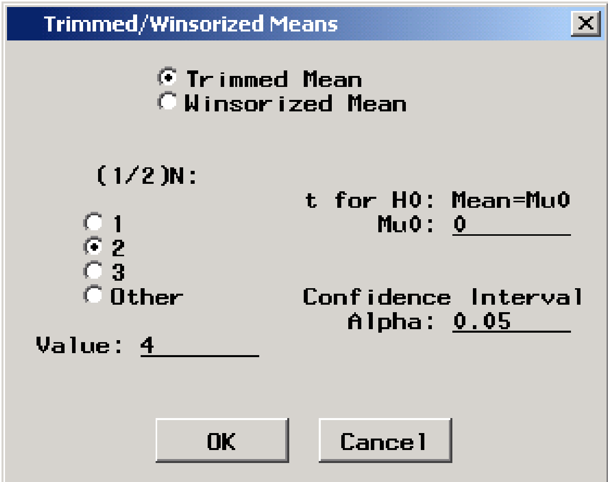
For H0: Variances are equal, F' = 5.32 DF = (11,13) Prob>F' = 0.0058

By the same token, for conducting analysis of variance in SAS, you can use PROC GLM (Procedure Generalized Linear Model) instead of PROC ANOVA when the data have unbalanced cells. However, the above adjusted t-test is only robust against the violation of equal variances. When multiple problems occur (welcome to the real world), such as non-normality, heterogeneous variances, and unequal sizes, the Type I error rate will inflate (Wilcox, 1998; Lix & Keselman, 1998).

To deal with the problem of multiple violations, robust methods such as trimmed means and Winsorized variances are recommended. In the former, outliers in both tails are simply omitted. In the latter, outliers are "pulled" towards the center of the distribution. For example, if the data vector is [1, 4, 4, 5, 5, 5, 6, 6, 10], the values "1" and "10" will be changed to "4" and "6," respectively. This method is based upon the Winsor's principle: "All observed distributions are Gaussian in the middle." Yuen (1974) suggested that to get the best of all methods, trimmed means and Winsorized variances should be used in conjunction with Welch's t-test (a form of robust t-test like Satterthwaite's).

SAS/Insight can compute both Winsorized and trimmed means by pointing and clicking (under the pull down menu “Table.” See Figure 1).

Figure 1. Winsorized and trimmed means in SAS/INSIGHT



In addition, PROC UNIVARIATE can provide the same option as well as robust measures of scale. By default, PROC UNIVARIATE does not return these statistics. The "ALL" option must be specified in the PROC statement to request the following results (see Figure 2).

Figure 2. Winsorized and trimmed means in PROC UNIVARIATE ALL

Trimmed Means

Percent Trimmed in Tail	Number Trimmed in Tail	Trimmed Mean	Std Error Trimmed Mean
25.00	9	3.777778	0.552935

Winsorized Means

Percent Winsorized in Tail	Number Winsorized in Tail	Winsorized Mean	Std Error Winsorized Mean
25.00	9	3.888889	0.561007

Robust Measures of Scale

Measure	Value	Estimate of Sigma
Interquartile Range	4.000000	2.965203
Gini's Mean Difference	2.834921	2.512383
MAD	2.000000	2.965200
Sn	2.385200	2.385200
Qn	2.221900	2.009759

Mallows and Tukey (1982) argued against the Winsor's principle. In their view, since this approach pays too much attention to the very center of the distribution, it is highly misleading. Instead, they recommended developing a way to describe the umbrae and penumbrae around the data. Tukey strongly endorsed using exploratory data analysis (EDA) techniques such as data transformation, which will be discussed in the next section.

In addition, Keselman and Zumbo (1997) found that the nonparametric approach has more power than the trimmed-mean approach does. Nevertheless, Wilcox (2001) asserted that the trimmed-mean approach is still desirable if 20 percent of the data are trimmed under non-normal distributions.

Regression analysis also requires several assumptions such as normally distributed residuals. When outliers are present, this assumption is violated. To rectify this situation, robust regression such as PROC LOESS can be used to downweight the influence of outliers. The weight range is from 0 to 1. Observations that are not extreme have the weight as "1" and thus are fully counted into the model. When the observations are outliers and produce large residuals, they are either totally ignored ("0" weight) or partially considered (low weight).

When data for ANOVA cannot meet the parametric assumptions, one can convert the grouping variables to dummy variables (1, 0) and run a robust regression procedure. As mentioned before, robust regression downweights extreme scores. When assumption violations occur due to extreme scores in one tail (skew distribution) or in two tails (wide dispersion, unequal variances), robust regression is able to compensate for the violations (Huynh & Finch, 2000).

Cliff (1996) was skeptical to the differential data-weighting of robust procedures. Instead he argued that data analysis should follow the principle of "one observation, one vote." Nevertheless, robust methods and conventional procedures should be used together when outliers are present. Two sets of results could be compared side by side in order to obtain a thorough picture of the data.

DATA TRANSFORMATION

Data transformation methods suggested by EDA are another alternative to counteract assumption violations (Behrens, 1997; Ferketich & Verran, 1994). Data transformation is also named data re-expression. The transformed data can be used in different ways. Because data transformation is tied to EDA, the data can be directly interpreted by EDA methods. Unlike classical procedures, the goal of EDA is to unveil the data pattern, and thus it is not necessary to make a probabilistic inference. Alternatively, the data can be further examined by classical methods if they meet parametric assumptions after the re-expression. Parametric analysis of transformed data is considered a better strategy than non-parametric analysis because the former appears to be more powerful than the latter (Rasmussen & Dunlap, 1991).

Data transformation happens in our everyday life. For example, converting US dollars into Canadian dollars, converting a GPA of 5-point scale to a GPA of 4-point scale. However, these examples belong to the linear transformation, by which the distribution of the data is not affected. In EDA, usually the non-linear transformation is used and thereby it changes the data pattern. Data re-expression is exploratory in nature because, prior to the transformation, the researcher never knows which re-expression approach can achieve desirable results. Cliff (1996) argued that data transformation confines the conclusion to the arbitrary version of the variables, indeed data transformation is hardly arbitrary. The following are common schemes of data re-expression:

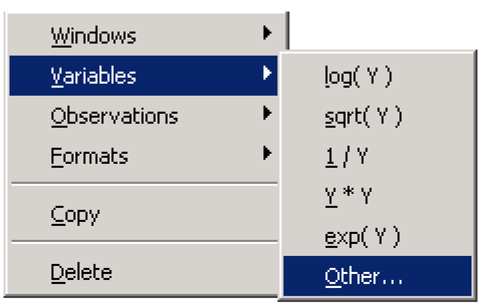
1. Normalize the distribution: Non-normal data violate the assumption of a parametric test and thus a transformation is advisable. It is a common misconception that converting raw scores to z-scores yields a normal distribution. Actually, the raw-to-z transformation is a linear transformation. The appropriate procedure should be natural log transformation or inverse probability transformation.

2. Stabilize the variances: Data with unequal variances are also detrimental to parametric tests. A typical example of variance stabilizing transformation is square root transformation: y* = sqrt(y).

3. Linearize the trend: Regression analysis requires the assumption of linearity. When the data show a curvilinear relationship, the researcher can either apply non-linear regression analysis or straighten the trend by linearizing transformation. A logarithmic transformation is a typical example of the latter.

In SAS/STAT data transformation can be performed by invoking different functions. However, SAS/INSIGHT is a better tool for this task because of its interactive nature. In SAS/INSIGHT, any variable can be highlighted and transformed by pre-determined transformation functions. Customized transformations can also be built by selecting "other…" (see Figure 3).

Figure 3. Data transformation in SAS/Insight



RESAMPLING

Resampling techniques such as randomization exact test, jackknife, and bootstrap are also viable alternatives (Diaconis & Efron, 1983; Edgington, 1995; Ludbrook & Dudley, 1998). Robust procedures recognize the threat of parametric assumption violations and make adjustments to work around the problem. Data re-expression converts data to ensure the validity of using parametric tests. Resampling is very different from the above remedies, for it is not under the framework of theoretical distributions imposed by classical parametric procedures.

Classical parametric tests compare observed statistics to theoretical sampling distributions. Resampling is a revolutionary methodology because it is departed from theoretical distributions. Rather the inference is based upon repeated sampling within the same empirical sample, and that is why this school is called resampling.

The resampling method is tied to the Monte Carlo simulation, in which researchers "make up" data and draw conclusions based on many possible scenarios. The name "Monte Carlo" came from an analogy to the gambling houses on the French Riviera. Many years ago when some gamblers studied how they could maximize the chances to win, they used simulations to check the occurrence of each case. Today Monte Carlo simulations are widely used by statisticians to study the "behaviors" of different statistical procedures. Nevertheless, there is a fundamental difference between Monte Carlo simulation and resampling. In the former data could be totally hypothetical while in the latter the simulation is still based upon some real data.

In older versions of the SAS system, resampling must be performed by using macros (SAS Institute, 1997). In the new version, PROC MULTTEST is available for bootstrapping and permutation tests. This procedure can exhaust every possible combination of outcomes, and thus an empirical distribution is built on the empirical sample data. Because one compares the observed statistics with the empirical distribution, the latter becomes the reference set.

Nonetheless, some methodologists are skeptical toward resampling for the following reasons:

1. Assumption: Fienberg said, "you're trying to get something for nothing. You use the same numbers over and over again until you get an answer that you can't get any other way. In order to do that, you have to assume something, and you may live to regret that hidden assumption later on" (cited in Peterson, 1991, p. 57).

2. Generalization: Some critics argue that resampling is based on one sample and therefore the generalization cannot go beyond that particular sample. One critic even went further to say, "I do wonder, though, why one would call this (resampling) inference?" (cited in Ludbrook & Dudley, 1998)

3. Bias and bad data: Bosch (2002) asserted that confidence intervals obtained by simple bootstrapping are always biased, though the bias decreases with sample size. If the sample comes from a normal population, the bias in the size of the confidence interval is at least n/(n-1), where n is the sample size. Some critics challenge that when the collected data are biased, resampling would just repeat and magnify the same mistake. Rodgers (1999) admitted that the potential magnification of unusual features of the sample is certainly one of the major threats to the validity of conclusions derived from resampling procedures.

4. Distributions are not identical: Cliff (1996) argued that randomization test of location comparisons, such as mean differences, are valid under the assumption that the distributions are identical. If they differ in spread or skewness, the Type I error rate can be inflated.

Taking the preceding potential problems of resampling into consideration, it is recommended that resampling and classical procedures should be used together.

CONCLUSION

Parametric tests undoubtedly have limitations. Unfortunately, in spite of repeated warnings, many researchers still proceed with those tests without implementing any remedy. They always assume that all tests are "ocean liners." It is hoped that the alternatives to parametric tests in the SAS system highlighted in this paper can be applied to rectify the situation. The preceding options are not mutually exclusive. Rather they can be used together to compliment each other and to verify the results.

NOTES

1. Today researchers very seldom use a single Likert scale as a variable. Instead, many items are combined as a composite score if Cronbach Alpha verifies that the items are internally consistent and factor analysis confirms that all items could be loaded into one single dimension. By using a composite score, some social scientists believe that the ordinal-scaled data based upon a Likert-scale could be converted into a form of pseudo-interval-scaled data. To be specific, if 50 five-point Likert-scaled items were totaled as a composite score, the possible range of data value would be from 1 to 250. In this case, a more extensive scale could form a wider distribution. Nonetheless, this argument is not universally accepted.

   The issue of the appropriateness of ordinal-scaled data in parametric tests was unsettled even in the eyes of Stevens (1951), the inventor of the four levels of measurement: "As a matter of fact, most of the scales used widely and effectively by psychologists are ordinal scales...there can be involved a kind of pragmatic sanction: in numerous instances it leads to fruitful results" (p.26). Based on the central limit theorem and Monte Carlo simulations, Baker, Hardyck, and Petrinovich (1966) and Borgatta and Bohrnstedt (1980) argued that for typical data, worrying about whether scales are ordinal or interval does not matter.

   Another argument against not using interval-based statistical techniques for ordinal data was suggested by Tukey (1986). In Tukey's view, this was a historically unfounded overreaction. In physics, before precise measurements were introduced, many physical measurements were only approximately interval scales. For example, temperature measurement was based on liquid-in-glass thermometers. But it is unreasonable not to use a t-test to compare two groups of such temperatures. Tukey argued that researchers painted themselves into a corner on such matters because they were too obsessed with "sanctification" by precision and certainty. If p-values or confidence intervals are to be sacred, they must be exact. In the practical world, when data values are transformed (e.g. transforming y to sqrt(y), or logy), the p values resulting from different expressions of data change. Thus, ordinal-scaled data should not be banned from entering the realm of parametric tests. For a review of the debate concerning ordinal- and interval-scaled data, please consult Velleman and Wilkinson (1993).

2. Harrell (1999) disagreed with Edgington: "Edgington's comment is off the mark in most cases. The efficiency of the Wilcoxon-Mann-Whitney test is 3/pi (0.96) with respect to the t-test IF THE DATA ARE NORMAL. If they are non-normal, the relative efficiency of the Wilcoxon test can be arbitrarily better than the t-test. Likewise, Spearman's correlation test is quite efficient relative to the Pearson r test if the data are bivariate normal. Where you lose efficiency with nonparametric methods is with estimation of absolute quantities, not with comparing groups or testing correlations. The sample median has efficiency of only 2/pi against the sample mean if the data are from a normal distribution."

REFERENCES

Baker, B. O., Hardyck, C. D., & Petrinovich, L. F. (1966). Weak measurement vs. strong statistics: An empirical critique of S. S. Stevens' proscriptions on statistics. Educational and Psychological Measurement, 26, 291-309.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. Psychological Methods, 2, 131-160.

Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of measurement: Once over again. Sociological Methods and Research, 9, 147-160.

Bosch, P. V. (2002 May). Personal communication.

Box, G. E. P. (1953). Non-normality and tests on variances. Biometrika, 40, 318-335.

Breckler, S. J. (1990). Application of covariance structure modeling in psychology: Cause for concern? Psychological Bulletin, 107, 260-273.

Cleveland, W. S. (1993). Visualizing data. Summit, NJ: Hobart Press.

Cliff, N. (1996). Answering ordinal questions with ordinal data using ordinal statistics. Multivariate Behavioral Research, 31, 331-350.

Diaconis, P., & Efron, P. (1983). Computer-intensive methods in statistics. Scientific American, May, 116-130.

Edgington, E. S. (1995). Randomization tests. New York : M. Dekker.

Ferketich, S., & Verran, J. (1994). An overview of data transformation. Research in Nursing and Health, 5, 393-404.

Fisher, R. A. (1935). The logic of inductive inference. Journal of the Royal Statistical Society, 98, 39-82.

Gibbons, J. D. (1993). Nonparametric statistics: An introduction. Newbury Park: Sage Publications.

Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology (3rd ed.). Boston, MA: Allyn and Bacon.

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 42, 237-288.

Harrell, F. E. (1999, December 8). Re: Disadvantages of nonparametric vs. parametric tests. Educational Statistics Discussion List (EDSTAT-L). [Online]. Available E-mail: edstat-l@jse.stat.ncsu.edu [1999, December 8].

Hodges J., & Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t test. Annals of Mathematical Statistics, 27, 324-335.

Huynh, H., & Finch, H. (2000 April). Robust/resistant statistical procedures with applications to multiple regression analysis and analysis of variance. AERA Professional Development and Training, New Orleans, LA.

Johnson, D. H. (1995). Statistical sirens: The allure of nonparametrics. Ecology, 76, 1998-2000.

Keselman, H. J., Huberty, C., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., & Keselman, J. C. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research, 68, 350-386.

Keselman, R. C., & Zumbo, B. (1997). Specialized tests for detecting treatment effects in the two-sample problem. Journal of Experimental Education, 65, 355-366.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasicity and nonnormality. Educational and Psychological Measurement, 58, 409-429.

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. American Statistician, 52, 127-132.

Mallows, C. L., & Tukey, J. W. (1982). An overview of techniques of data analysis, emphasizing its exploratory aspects. In J. T. de Oliveira & B. Epstein (Eds.), Some recent advances in statistics (pp. 111-172). London: Academic Press.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Peterson, I. (1991, July 27). Pick a sample. Science News, 140, 56-58.

Rasmussen, J. L., & Dunlap, W. P. (1991). Dealing with nonnormal data: Parametric analysis of transformed data vs. nonparametric analysis. Educational & Psychological Measurement, 51, 809-820.

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sample taxonomy. Multivariate Behavioral Research, 34, 441-456.

SAS Institute (1997). Subsetting and resampling in SAS [On-line]. Available: http://www.utexas.edu/acits/docs/stat56.html.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: John Wiley.

Tanizaki, H. (1997). Power comparison of non-parametric tests: Small-sample properties from Monte Carlo experiments. Journal of Applied Statistics, 24, 603-632.

Thompson, M. S., Green, S. B., Stockford, S. M., Yu, C. H., Chen, Y., & Lo, W. (2002 April). The .05 level: The probability that the independent-samples t test should be applied? Paper presented at the Annual Meeting of the American Education Researcher Association, New Orleans, LA.

Tukey, J. W. (1986). The collected works of John W. Tukey (Volume IV): Philosophy and principles of data analysis 1965-1986. Monterey, CA: Wadsworth & Brooks/Cole.

Velleman, P. F., & Wilkinson, P. F. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. The American Statistician, 47, 65-72.

Wilcox, R. (2001). Fundamentals of modern statistical methods: Substantially improving power and accuracy. New York: Springer Verlag .

Yuen , K. K. (1974). The two-sample trimmed t for unequal population variances. Biometrika, 61, 165-170.

Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. Journal of Experimental Education, 67, 55-68.

CONTACT INFORMATION

Chong Ho Yu, Ph.D., CCNA, MCSE, CNE
Psychometrician
PO Box 612
Tempe AZ 85280
USA
602-778-2722
Email: asumain@yahoo.com.hk
Website: http://seamonkey.ed.asu.edu/~alex