REVIEWS

An Introduction to Non-parametric Statistics for Health Scientists

Christy C Tomkins MSc, CK
Faculty of Rehabilitation Medicine

Correspondence to: Christy Tomkins. 3-50 Corbett Hall, University of Alberta, Edmonton, AB, T6G 2G4. Phone (780)432-3982 E-mail: ctomkins@ualberta.ca

ABSTRACT

Data from research in the health sciences often violates one, if not all of the assumptions underlying the use of traditional parametric statistics. Non-parametric statistics are another group of tests for statistical inference, which do not make strict assumptions about the population from which the data have been sampled, and may be used for studies with small sample sizes, nominal or ordinal level data, and non-normally distributed variables. Non-parametric techniques are widely applicable to research in the health sciences. The challenge appears to be convincing researchers to adopt these techniques, despite misconceptions and limited exposure. The following review will provide a brief introduction to non-parametric statistics, as they apply to research conducted in the field of health sciences.

KEY WORDS: non-parametric statistics, health sciences, sample size, measurement, research

Many introductory courses in statistics tend to focus on what are called *parametric* statistics. These techniques are termed parametric because they focus on specific parameters of the population, usually the mean and variance (1). In order to utilize these techniques, a number of assumptions regarding the nature of population from which the data are drawn must be met (Table 1). One might note that typical health science research often violates one, if not all of these parametric assumptions. The solution to this problem in many cases is another group of tests for statistical inference, which do not make strict assumptions about the population from which the data have been sampled: *non-parametric statistics* (1). Although non-parametric techniques do not require the stringent assumptions associated with their parametric counterparts, this does not imply that they are 'assumption free' (Table 2).

There are three major parametric assumptions, which are, and will continue to be routinely violated by research in the health sciences: level of measurement, sample size, and normal distribution of the dependent variable. The following sections will discuss these assumptions, and elucidate why much of the data procured in health science research violate these assumptions, thus implicating the use of non-parametric techniques.

Level of Measurement

When deciding which statistical test to use, it is important to identify the level of measurement associated with the dependent variable of interest. Generally, for the use of a parametric test, a minimum of interval level measurement is required. Nonparametric techniques can be used with all levels of measurement, and are most frequently associated with nominal and ordinal level data.

Nominal data

The first level of measurement is nominal, or categorical. Nominal scales are usually composed of two mutually exclusive named categories with no implied ordering: yes or no, male or female. Data are placed in one of the categories, and the numbers in each category are counted (also known

Table 1. Assumptions to be met for the use of parametric tests

- Normal distribution of the dependent variable
- A certain level of measurement: Interval data
- Adequate sample size (>30 recommended per group)
- An independence of observations, except with paired data
- Observations for the dependent variable have been randomly drawn
- Equal variance among sample populations
- Hypotheses usually made about numerical values, especially the mean

Adapted from: Pett MA. Nonparametric Statistics for Health Care Research. London, Thousand Oaks, New Delhi: Sage Publications, 1997.

Table 2. Characteristics common to most non-parametric techniques (15)

- Fewer assumptions regarding the population distribution
- Sample sizes are often less stringent
- Measurement level may be nominal or ordinal
- Independence of randomly selected observations, except when paired
- Primary focus is on the rank ordering or frequencies of data
- Hypotheses are posed regarding ranks, medians, or frequencies of data

Adapted from:Pett MA. Nonparametric Statistics for Health Care Research. London, Thousand Oaks, New Delhi: Sage Publications, 1997.

Table 3. Example of an item using a nominal level measurement scale

1. Does your back problem affect your employment status?

   ❑ Yes ❑ No

2. Are you limited in how many minutes you are able to walk continuously with or without support (i.e. cane)?

   ❑ No ❑ Yes

Table 4. Example of an item using an ordinal level measurement scale

Section 4 – Walking

1 Pain does not prevent me walking any distance
2 Pain prevents me from walking more than 2 kilometres
3 Pain prevents me from walking more than 1 kilometre
4 Pain prevents me from walking more than 500 meters
5 I can only walk using a stick or crutches
6 I am in bed most of the time

as frequencies). The key to nominal level measurement is that there are no numerical values assigned to the variables. Given that no ordering or meaningful numerical distances between numbers exist in nominal measurement, we cannot obtain the coveted 'normal distribution' of the dependent variable (1). Descriptive research in the health sciences would make use of the nominal scale often when collecting demographic data about target populations (i.e. pain present or not present, ambulatory or not ambulatory, agree or disagree). Examples of items representing nominal level measurement are presented in Table 3.

Ordinal data

The second level of measurement, which is also frequently associated with non-parametric statistics, is the ordinal scale (also known as rank-order). Ordinal level measurement gives us a quantitative 'order' of variables, in mutually exclusive categories, but no indication as to the value of the differences between the positions (squash ladders, army ranks) (1). As such, the difference between positions in the ordered scale cannot be assumed to be equal. Examples of ordinal scales in health science research include pain scales, stress scales, and functional scales. One could estimate that someone with a score of 5 is in more pain, more stressed, or more functional than someone with a score of 3, but not by *how much.* There are a number of non-parametric techniques available to test hypotheses about differences between groups and relationships among variables, as well as descriptive statistics relying on rank ordering. Table 4 provides an example of an ordinal level item from the Oswestry Disability Index (2).

Interval and ratio data

Interval level data is usually a minimum requirement for the use of parametric techniques. This type of data is also ordered into mutually exclusive categories, but in this case the divisions between categories are equidistant. The only difference between interval data and ratio data, is the presence of a meaningful zero point. In interval level measurement, zero does not represent the absence of value. As such, you cannot say that one point is say, two times larger than another. For example, 100 degrees Celsius is not two times hotter than 50 degrees because zero does not represent the complete absence of heat.

Ratio is the highest level of measurement and provides the most information. The level of measurement is characterized by equal intervals between variables, and a meaningful zero point. Examples of ratio level measurement include weight, blood pressure, and force (1).

It is important to note that in health science research we often use multi-item scales, with individual items being either nominal or ordinal. Table 5 provides an example of such items from the CES-D depression scale (3).

The item responses (0-3) are summed to produce a scale with interval-level properties and a larger range of scores (in this case the range would be 0-15) (1). It is possible to use parametric statistics, which consider the mean and variance in such cases.

Sample size

Adequate sample size is another of the assumptions underlying parametric tests. It is not uncommon to see small sample sizes (i.e. n=5), or case studies (one subject) in health science literature. Health scientists frequently work with small groups of individuals, low incidence conditions, convenience samples, and limited funding. Thus, the assumption of large sample size is often violated by such studies using parametric statistical techniques.

The sample size required for a study has implications for both choices of statistical techniques and resulting power (1). It has been shown that sample size is directly related to researchers' ability to correctly reject a null hypothesis (power) (4;5). As such, small sample sizes often reduce power and increase the chance of a type II error. It has been found that by using non-parametric techniques with small sample sizes, it is possible to gain adequate power. However, there does not seem to be a consensus among statisticians regarding what constitutes a small sample size. Siegel and Castellan (6) argue that if the sample size is very small, there may be no alternative to using a non-parametric statistical test (6), but the value of 'very small' is not delineated. It has been suggested by Wampold et al. (7), that the issue of sample size is closely related to the distribution of the dependent variable, given that as sample size increases, the sampling distribution approaches normal (n>100). It was suggested that if the distribution of the dependent variable resembles closely a normal distribution then moderate sample sizes (5 or 10) will be sufficient to say that the sampling distribution of the mean is approximately normal. For other distributions, 30 observations might be required (7).

As with many facets of research, decisions regarding statistical technique choice are not cut and dry, but require us to 'get dirty with the data'. In light of the dearth of consensus regarding

Continued on page 24

Human Rights Photo Spread

A young girl in Botswana who dreams of becoming a teacher is taken out of school to help her mother take care of their family… A well educated Indian man faces discrimination because his last name indicates a lower class status… A family mourns the loss of their son, who was killed in war… A Ukrainian family forced to live in an internment camp simply because of their origin… Another child dies from a treatable infection because their family had no access to medical care…

The Universal Declaration of Human Rights, Article 1 states: *All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.* It goes on to state that everyone has the right to life, liberty and the security of person and that no one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment. The declaration exists to protect the rights of every individual from the day they are born until the day they die, no matter who they are, where they are or what they believe in.

What we can't see, can't hurt us. In reality, there are many parts of the world where not only are individual human rights ignored, but they are violated. It is all too easy for these violations to go unrecognized and even if they are acknowledged, action is not easily elicited. The following pictures were taken by U of A medical students and illustrate what they have seen around the world, and their reflections on the importance of valuing human rights.

Take a moment to see for yourself the reality of the situation. Take a moment to think about what you could do to make a difference. Take a moment to make a difference.

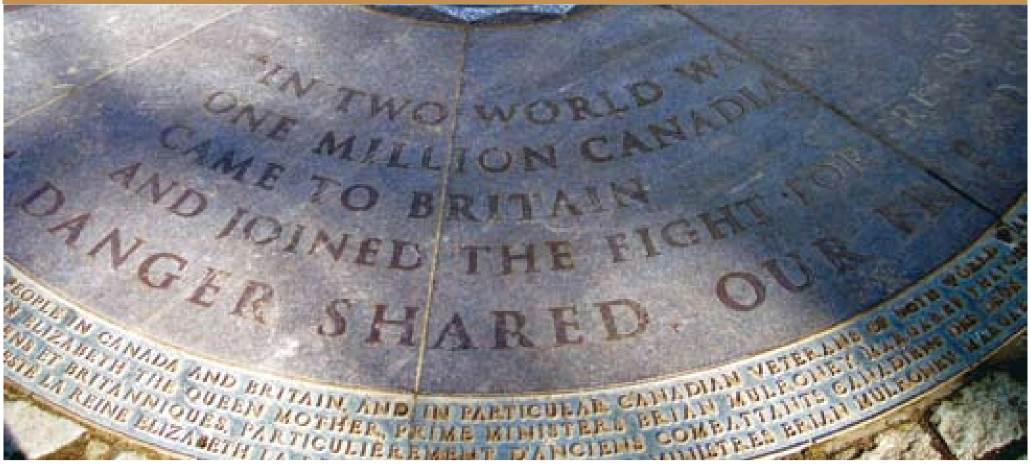
Shirmee Doshi, Medicine, Class of 2009

"Article 26 of the Universal Declaration of Human Rights states that everyone has the right to education and that it shall be free, at least in the elementary and fundamental stages. As in many parts of the world, Bénin educational statistics indicate that women lag behind their male counterparts in the successful completion of elementary and secondary school. This photo depicts children at an early stage in their education, when the gender balance is equal. To create a balanced and diverse workforce, equal access to education must be strived for."

Parakou, Bénin, Western Africa By: Graeme Brassard, Class of 2006

"The monument represents the drastic actions that are occasionally required to stop invasions and threats to basic human rights, such as housing, safety and health. This also demonstrates the importance that Canadians and individuals around the world place on human rights, and the extremes they will go to in their commitment to protect the rights of all individuals."

London, England By: Kelvin Leung, Class of 2008



"Patang Baazi is an annual kite flying festival, where families climb to their rooftops and spend the day flying kites, cutting down their competition and enjoying a day of leisure. It is the one day in India where there are no boundaries of castes, creed, color, religion, nationality or language. For one day, everyone is equal and the fighting and discrimination stops. But only for one day."

Baroda, India By: Shirmee Doshi, Class of 2009

"Safety, health and education take a back seat to the constant struggle for survival and search for family income. Working long hours for little pay, becoming involved in illegal acts or being maliciously disfigured to make better beggars were not uncommon situations. It's heartbreaking to convey, but for many children of India childhood is lost… except for those brief moments in the sand."

Chowpatty Beach, Mumbai, India By: Laine Racher, Class of 2008

"We often associate a lack of human rights with places overseas and far away; this picture was taken in our own backyard, and serves as a reminder that human rights are still an issue here and now in Canada."

Banff, Canada By: Clark Schommer, Class of 2009

"Many families from Burma swim the Moei River to gain illegal entry to Thailand in order to access medical treatment. Too often I was witness to children not being able to enjoy their childhood because of the responsibilities imposed on them. The family's poverty requires many children to become the caretakers of the siblings admitted to the local clinic - after seeing such demands placed on children, I was delighted to see these boys enjoying themselves at play."

Mae Tao Clinic, Mae Sot, Thailand
By: Geneil Campbell, Class of 2006

"Universal Declaration of Human Rights, Article 4: No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms. This picture is of a little Puruhae (indigenous Andean) boy. This little boy's parents are the first generation to live freely and own their own land. They are finally making enough money to provide an education for their son and to achieve a better quality of life for their family."

Guargualla, Ecuador By:
Nipunie Rajapakse, Class of 2009

Continued from page 21

Table 5. Example of ordinal level measurement

During the past 1 week:	Rarely or none of the time (Less than a day)	Some or a little of the time (1-2 days)	Occasionally or a moderate amount of the time (3-4 days)	Most or all of the time (5-7 days)
1. I was bothered by things that usually don't bother me	0	1	2	3

what constitutes a small sample, Pett has suggested that the choice of parametric or non-parametric tests just 'depends' (1). It depends on sample size, level of measurement, the researcher's knowledge of the variables' distribution in the population, and the shape of the distribution of the variable of interest. If in doubt, try using both parametric and non-parametric techniques.

Normality

According to Pett, in choosing a test we must consider the shape of the distribution of the variable of interest. In order to use a parametric test, we must assume a normal distribution of the dependent variable (1). However, in real research situations things do not come packaged with labels detailing the characteristics of the population of origin (8). Sometimes it is feasible to base assumptions of population distributions on empirical evidence, or past experience. However, often sample sizes are too small, or experience too limited to make any reasonable assumptions about the population parameters (1). Generally in practice, one is only able to say that a sample appears to come from say, a skewed, very peaked, or very flat population (9). Even when one has a precise measurement (ratio scale), it may be irrational to assume a normal distribution, because this implies a certain degree of symmetry and spread (9). Non-parametric statistics were designed to be used when we know nothing about the distribution of the variable of interest. Thus, we can apply non-parametric techniques to data from which the variable of interest does not belong to any specified distribution (i.e. normal distribution).

Although there are many variables in existence that are normally distributed, such as weight, height and strength, this is not true of all variables in health science (10). The incidence of rare disease and low prevalence conditions are both non-normally distributed populations. However, it seems that most researchers using parametric statistics often just 'assume' normality. Micceri et al. (11) state that the naïve assumption of normality appears to characterize research in many fields. However, empirical studies have documented non-normal distributions in literature from a variety of fields (11). Micceri et al. (11) investigated the distribution in 440 large sample achievement and psychometric measures. It was found that *all* of the samples were significantly *non-normal* ($p<0.01$). It was concluded that the underlying tenets of normal-assuming statistics appeared to be fallacious for the commonly used data in these samples (11). It is likely that if a similar study, investigating the nature of the distributions of data were to be conducted with some of the measures commonly used in health science research, a similar result would ensue, given that not all variables are normally distributed.

When and why to use non-parametric techniques

It is apparent that there are a number of factors involved in choosing whether or not to use a non-parametric test, including level of measurement, sample size and sample distribution. Table 6 summarizes a number of situations that would implicate the use of a non-parametric technique. When the choice of statistical technique for a set of data is not clear, there is no harm in analyzing the data both ways: using both parametric and non-parametric tests. For each of the main parametric techniques there is a non-parametric counterpart; Experiment with the data to determine which test provides the best power, and the greatest level of significance.

There are a number of advantages to using non-parametric techniques in health science research (Table 7). The most important of these advantages are the generality and wide scope of non-parametric techniques. The lack of stringent assumptions associated with non-parametric tests implies that there is little probability of violating assumptions, which implies robustness. The application of non-parametric tests in health science research is wide, given that they can be applied to constructs for which it is impossible to obtain quantitative measures (descriptive studies), as well as to small sample sizes.

Misconceptions about non-parametric statistics

The lack of use of non-parametric techniques is owing to a series of common misconceptions about this branch of statistics. Non-parametric statistics have long taken the back seat to parametric statistics, often being portrayed as inferior in practice and teaching. It has been suggested that researchers are hesitant to use these techniques, due to fears that peer reviewers may not be completely familiar with these statistics, and therefore unable to properly interpret, and review the results (12). This opinion could be owing to a widespread case of limited exposure of research clinicians to this type of statistics. Non-parametric techniques are often left out of basic statistics courses, and relegated to the last chapter of texts (12), making them seem less important, while reinforcing the focus on parametric statistics.

Another common misconception concerning non-parametric statistics is that they are restricted in their application (12). It is thought that there are only a limited number of simple designs that can be analyzed using these techniques. However, there are non-parametric techniques which span from simple 2-group analysis, to complex structural equation modelling (12). Basically, for any parametric test, there is a non-parametric equivalent that would be equally, or in some cases, more appropriate for use.

Table 6. Conditions when it is appropriate to use a non-parametric test

- Nominal or ordinal level of measurement
- Small sample sizes
- Non-normal distribution of dependent variable
- Unequal variances across groups
- Data with notable outliers

Table 7. Advantages of non-parametric statistics for use in health science research

- Methods quick and easy to apply
- Theory fairly simple
- Assumptions for tests easily satisfied
- Accommodate unusual or irregular sample distributions
- Basic data need not be actual measurements
- Use with small sample sizes
- Inherently robust due to lack of stringent assumptions
- Process of collecting data may conserve time and funds
- Often offer a selection of interchangeable methods
- Can be used with samples made up of observations from several different populations

Adapted from: Pett MA. Nonparametric Statistics for Health Care Research. London, Thousand Oaks, New Delhi: Sage Publications, 1997.

Are parametric tests more powerful?

The primary barrier to use of non-parametric tests is the misconception that they are less powerful than their parametric counterparts (power is the ability to correctly reject the null hypothesis). It has been suggested that parametric tests are almost always more powerful than non-parametric tests (13). This assertion is still made in current textbooks, such as the Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation (2003) (10). These assertions are often made with no references to support them, suggesting that this falls into the realm of 'common knowledge' (14). Evidence to support this is not abundant, nor conclusive (14). Rather, on closer examination, it is found that parametric tests are more powerful than nonparametric tests *only if* all of the assumptions underlying the parametric test are met (1;8;10;12;14;15).

Pierce (15) suggests that unless it has been determined that the data do comply with all of the restrictions imposed by the parametric test, the greater power of the parametric test is irrelevant. This is because 'the purpose of applied statistics is to delineate and justify the inferences that can be made within the limits of existing knowledge - that purpose is defeated if the knowledge assumed is beyond that actually possessed' (15). Thus, the power advantage of the parametric test *does not hold* when the assumptions of the parametric test are not met, when the data are in ranks, or when the nonparametric test is used with interval or ratio data (12).

When comparison studies have been made between parametric and non-parametric tests, the non-parametric tests are frequently as powerful as parametric, especially with smaller sample sizes (16;17). Blair et al. (14) compared the power of the paired sample t-test (a common parametric test), to the Wilcoxon signed-ranks test (non-parametric), under various population shapes and sample sizes (n=10, 25, 50), using a simple pre-post test design (14). It was found that in some situations the t-test was more powerful than the Wilcoxon. However, the Wilcoxon test was found to be the more powerful test in a greater number of situations (certain population shapes and sample sizes), especially when sample sizes were small (14). In addition, the power advantage of the Wilcoxon test often increased with larger sample sizes, suggesting that non-parametric techniques need not be limited to studies with small sample sizes (14). It was concluded that insofar as these two statistics are concerned, the often-repeated claim that parametric tests are more powerful than non-parametric test is not justified (14). Generally, the rationale for using the t-test over the Wilcoxon test is that the parametric tests are more powerful under the assumption of normality. However, it was shown in this study that even under normal theory, there was little to gain, in terms of power by using the t-test as opposed to the Wilcoxon (14). It was suggested by Blair (14) that 'it is difficult to justify the use of a t-test in situations where the shape of the sampled population is unknown on the basis that a power advantage will be gained if the populations does *happen* to be normal' (14). Blair concluded by saying that 'although there were only two tests compared here, it should be viewed as part of a small but growing body of evidence that is seriously challenging the traditional views of nonparametric statistics' (14). This study demonstrated that the use of non-parametric techniques is implicated whenever there is doubt regarding the fulfillment of parametric assumptions, such as normality or sample size.

Which non-parametric test should we use?

The most common non-parametric tests can be found in Table 8. Please refer to the following statistical texts for the derivation and calculation of these statistics, as this is beyond the scope or intention of this paper: Nonparametric Statistics for the Behavioural Science (Siegel Sand Castellan NJ, 1988) (6), Applied Nonparametric Statistical Methods (Sprent P and Smeeton NC, 2001) (9), Nonparametric Statistical Inference (Gibbons JD, 1985) (8), Nonparametrics: Statistical Methods Based On Ranks (Lehmann EL, 1975) (18), Practical Nonparametric Statistics (Conover WJ, 1980) (19), Fundamentals of Nonparametric Statistics (Pierce A, 1970) (15), and Essentials of Research Methods in Health, Physical Education, Exercise Science and Recreation (Berg KE and Latin RW, 2003) (10).

Conclusion

The proper application of statistical techniques is essential for the effective and accurate interpretation, and presentation of data in health science research. Thus, it is important that researchers become aware of the statistical techniques available and know when to apply them. It is apparent that non-parametric techniques are widely applicable to research in the health sciences. The challenge appears to be convincing researchers to adopt these techniques, despite misconceptions and limited exposure. The expansion of knowledge and advance of science depends on ingenuity. For knowledge to advance in the field of health sciences, researchers must be willing to try new techniques and challenge the traditional views of non-parametric statistics, realizing that these techniques are likely

Table 8. Common nonparametric statistics

Purpose of Test	Level of measurement	Nonparametric statistic
"Goodness of Fit"	Nominal	• Chi-Square-Goodness-of-Fit Test • The Binomial Test
	Ordinal/Interval	• Kolmogorov Smirnov One-Sample and Two-Sample Tests
Related Samples: Pre-test-post-test measures for a single sample (2 measures)	Nominal	• The McNemar Test
	Ordinal/Interval	• The Sign Test • Wilcoxon Signed Ranks Test
Related Samples: Pre-test-post-tests measures for a single or matched sample (>2 measures)	Nominal	• Cochran's Q Test
	Ordinal/Interval	• The Friedman Test
Tests for Two Independent Groups	Nominal	• The Fisher Exact Test (dichotomous variable) • Chi-square Test for Two Independent Samples
	Ordinal/Interval	• The Wilcoxon-Mann-Whitney U Test
Assessing Differences Among Several Independent Groups	Nominal	• Chi-square Test for k Independent Samples • The Mantel-Haenszel Chi-Square Test for Trends
	Ordinal/Interval	• The Median Test • Kruskal Wallis-One Way ANOVA by Ranks
Tests of Association Between Variables	Nominal	• Phi Coefficient (dichotomous variable) • Cramér's V Coefficient • The Kappa Coefficient
	Ordinal/Interval	• Point Biserial Correlation • Spearman's Rank Order Correlation Coefficient • Kendall's Tau Coefficient

Adapted from: Pett MA. Nonparametric Statistics for Health Care Research. London, Thousand Oaks, New Delhi: Sage Publications, 1997.

best suited for a substantial proportion of the current research in this field.

References

1. Pett MA. Nonparametric Statistics for Health Care Research. London, Thousand Oaks, New Delhi: Sage Publications, 1997.
2. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;2940-52.
3. Cole JC, Rabin A.S., Smith T.L. Development and Validation of a Rasch-Derived CES-D Short Form. Psychological Assessment 2004;360-72.
4. Cohen J. Statistical Power Analysis for the Behavioural Sciences. Hillsdale, NJ: Lawrence Erlbaum, 1988.
5. Kraemer HC, Thiemann S. How Many Subjects? Newbury Park, CA: Sage, 1987.
6. Siegel S, Castellan NJ. Nonparametric Statistics for the Behavioural Sciences. New York: McGraw-Hill Book Company, 1988.
7. Wampold BE, Drew CJ. Theory and application of statistics. New York: McGraw-Hill, 1990.
8. Gibbons JD. Nonparametric Statistical Inference. New York: Marcel Dekker Inc., 1971.
9. Sprent P, Smeeton NC. Applied Nonparametric Statistical Methods. New York: Champan and Hall/CRC, 2001.
10. Berg KE, Latin RW. Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation. Philadelphia, Baltimore, New York, London, Buenos Aires, Hong Kong, Sydney, Tokyo: Lippincott Williams and Wilkins, 2004.
11. Micceri T. The Unicorn, The Normal Curve, and Other Improbable Creatures. Psychological Bulletin 1989;156-66.
12. Hunter MA, May RB. Some myths concerning parametric and nonparametric tests. Canadian Psychology 1993;384.
13. Kerlinger FN. Foundations of Behavioural Research. New York: Holt, Rinehart and Winston, 1964.
14. Blair RC, Higgins JJ. Comparison of teh Power of the Paired Samples t Test to That of Wilcoxon's Signed-Ranks Test Under Various Population Shapes. Psychological Bulletin 1985;119-28.
15. Pierce A. Fundamentals of Nonparametric Statistics. Belmont, California: Dickenson Publishing Company, 1970.
16. Blair RC, Higgins JJ. A comparison of the power of Wilcoxon's rank-sum statistic to that of Student's t statistic under various non-normal distributions. Journal of Educational Statistics 1980;309-35.
17. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician 1981;124-8.
18. Lehmann EL. Nonparametrics: Statistical Methods Based on Ranks. San Francisco, New York: Holden-Day Inc. and McGraw-Hill International Book Company, 1975.
19. Conover WJ. Practical Nonparametric Statistics. New York, Chichester, Brisbane, Toronto: John Wiley and Sons, 1980. ■