

## Nonparametric statistical methods

Sijtsma, K.; Emons, W.H.M.

*Published in:*  
International encyclopedia of education

*Document version:*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2010

[Link to publication](#)

*Citation for published version (APA):*  
Sijtsma, K., & Emons, W. H. M. (2010). Nonparametric statistical methods. In B. McGaw, E. Baker, & P. P. Peterson (Eds.), *International encyclopedia of education* (pp. 347-353). Elsevier.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

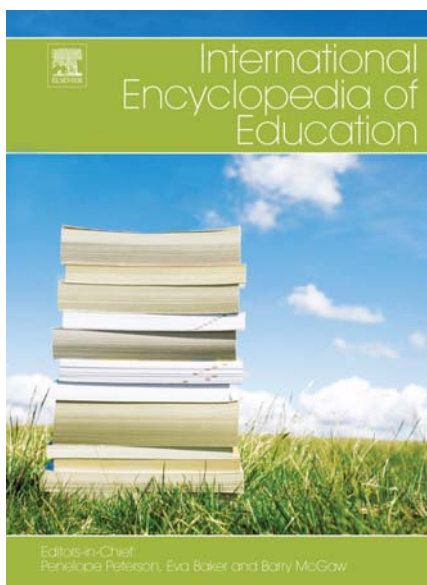
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Provided for non-commercial research and educational use.  
Not for reproduction, distribution or commercial use.**

This article was originally published in the *International Encyclopedia of Education* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

Sijtsma K and Emons W H M (2010), Nonparametric Statistical Methods. In: Penelope Peterson, Eva Baker, Barry McGaw, (Editors), *International Encyclopedia of Education*. volume 7, pp. 347-353. Oxford: Elsevier.

# Nonparametric Statistical Methods

K Sijtsma and W H M Emons, Tilburg University, Tilburg, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

## Glossary

**Nonparametric statistical methods** – The statistical methods that do not assume a particular distribution, such as the normal, from which the data are sampled.

## Introduction

Nonparametric statistical methods are used in situations in which it is unreasonable to assume that the sample was drawn from a population distribution with a particular parametric shape, such as the normal distribution. For example, when the sample size is small and the distribution of the observations is skewed, statistical methods based on the assumption of a normal distribution may be inappropriate. Nonparametric statistical methods do not have *a priori* assumptions about the population distribution and, therefore, are sometimes referred to as distribution-free methods.

Nonparametric methods have several properties that are worth mentioning. First, they are suitable for categorical data, which are common in educational research. Categorical variables have either nominal or ordinal measurement level, but may also represent counts of particular events. Examples of nominal variables include the kinds of arithmetic errors made by students, and the assignment of each student to one of three remedial teaching methods of interest. In statistical analysis, such groups may be identified by means of numbers 1, 2, and 3, which only serve to distinguish them. An example of an ordinal variable is the teacher's rank ordering of students with respect to perceived need of remedial teaching. An example of a count is the frequency of students in each remedial teaching program who receive help.

Nonparametric methods may also be in order for interval-level variables, such as students' motivation, for which the population distribution is unknown. The consequence of using nonparametric methods is that the test scores are treated as ordinal rather than interval variables. By doing this, information is lost about the distances between scores and the power of statistical tests is reduced. Researchers often consider this unacceptable, and analyze

their data as if the sampled scores on a variable stem from a known distribution.

Second, nonparametric methods are appreciated for their weak assumptions. Thus, instead of assuming distributions such as the normal, exact distributions are derived for particular statistics of interest. For example, a nominal variable may follow a multinomial distribution of which the category probabilities are estimated from the data. Other assumptions may be relaxed or completely dropped. For example, instead of assuming a linear relationship between numerical variables as in linear regression, relationships may be estimated from the data, as in kernel smoothing (Fox, 2000a; Ramsay and Silverman, 1997). Weak statistical models are important because they do not make assumptions beyond the level of knowledge expressed in many theories in educational and other social and behavioral sciences.

Third, it is sometimes said that nonparametric statistical models are convenient because they are easy to use. Some restraint may be in order here, because nonparametric methods are often based on complex mathematical considerations that may not be as easy to grasp as those underlying parametric methods. For example, several nonparametric methods are based on combinatorial math, which is notorious for running into badly manageable computations as sample size increases or the numbers of variables increases, and all possible patterns of scores have to be taken into account.

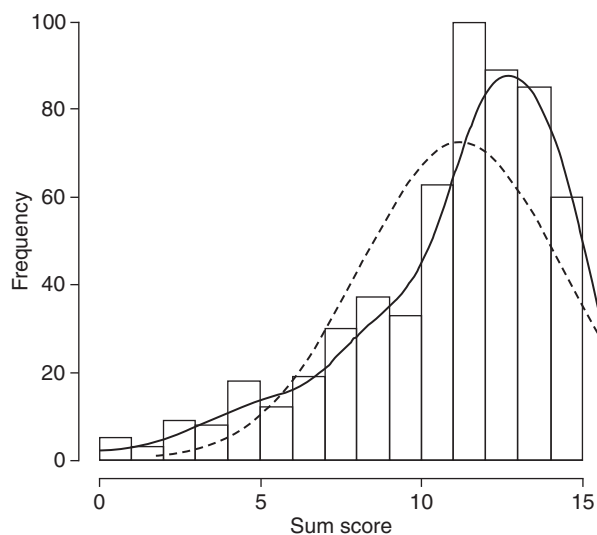
Fourth, the results of nonparametric statistical methods are sometimes not much different from those obtained by means of parametric methods that have been applied even when the assumptions on which they are based were violated. Thus, sometimes parametric statistical methods are robust against the violation of the normality assumption and the use of nonparametric counterparts, which may be more appropriate from a mathematical point of view, becomes less salient. This is true, for example, for the parametric Student's *t*-test. However, for other methods, such as regression analysis, it has been shown repeatedly that the estimation of a relationship from the data may lead to different and interesting results that would have been obscured when a particular parametric function, such as the linear or the logistic, had been fitted. Thus, there is room for nonparametric statistical methods, which are also illustrated by some examples provided in the article.

## Examples of Nonparametric Statistical Methods

The goal of statistical analysis is to estimate properties of interest from the data, such as the distribution of a variable or the difference between groups with respect to the means on a variable, and to test hypotheses about interesting research questions. First, we review a few well-known nonparametric statistical tests from the myriad of test procedures (e.g., Siegel, 1956; Siegel and Castellan, 1988; Wasserman, 2004, 2006) based on known, exact sampling distributions, and then we discuss methods for obtaining sampling distributions and probabilities of exceedence when exact sampling distributions are unknown.

### Estimating Distributions and Drawing Inferences

An important step in data analysis is to inspect the distribution of the observations. Rather than assuming a normal distribution and estimating the mean and variance, one can also estimate the complete distribution from the data. An example of a nonparametric method is the simple histogram that estimates the population distribution directly from the data (Figure 1). As the sample has a limited number of observations, the histogram is discrete and jagged. Thus, sometimes it may be convenient to smooth the histogram. Kernel smoothing produces the result shown by the solid curve in Figure 1. Smoothing has the effect of bringing out the salient features of the distribution at the expense of irregularities that probably are due to sampling error. For example, the solid smooth



**Figure 1** Example of an estimated distribution of sum scores (e.g., the number-correct score on an educational test) by means of a histogram, kernel smoothing (solid curve), and a normal approximation (dashed curve).

curve suggests that the distribution is skewed to the left and that the frequency of score 12 may be too large due to sampling error.

The smooth dashed curve in Figure 1 shows the normal approximation to the histogram. A glance at the graph shows that the normal seems to overestimate the lower-score frequencies and underestimate the higher-score frequencies. The Shapiro–Wilk test, which is a well-known nonparametric test for evaluating whether the observations deviate from the normal curve, yields a value equal to 0.894 ( $P < 0.000$ ); thus, the hypothesis of normality is rejected. The Kolmogorov–Smirnov test is a more general, often-used nonparametric method that can be used to test whether the data come from a hypothesized distribution, such as the normal. Often, it has less power than the Shapiro–Wilk test to detect violations of normality, but for the data in Figure 1 the value of the test statistic is 0.183 ( $P < 0.000$ ). Again, normality is rejected. Neither test makes assumptions about the population distribution of the data.

Researchers may wish to know whether the distributions of an outcome variable are the same across groups, for example, as in a control-group study on the effect of a teaching program. When the variable is ordinal, it does not make sense to assume that it follows a normal distribution and compare the means using a parametric  $t$ -test. Such a test is also inappropriate when an interval variable does not have a normal distribution. Alternatively, the nonparametric Mann–Whitney  $U$  test, also known as the Wilcoxon rank-sum test, is a good candidate for testing the null hypothesis that two independent samples come from the same population against the alternative that the samples come from two different population distributions which are identical in shape but different in location.

The Mann–Whitney  $U$  test is based on the common rank ordering according to ascending magnitude of all observations from two samples, say  $A$  with size  $N_1$  and  $B$  with size  $N_2$ , and counts for each observation from  $B$  by how many observations from  $A$  it is preceded; test statistic  $U$  equals the sum of these counts (for simplicity, we ignore the possibility of ties). When the two distributions are completely separated, each observation from  $A$  either precedes each observation from  $B$ , which results in maximum  $U = N_1 N_2$ , or is preceded by each observation from  $B$ , which results in minimum  $U = 0$ . Thus, high and low values of  $U$  indicate different distributions, and intermediate values indicate largely overlapping distributions. For small sample sizes (i.e., the size of the larger sample does not exceed 20), the probabilities of exceedence can be read from tables especially prepared for this purpose. When the larger sample exceeds size 20,  $U$  has been shown to approach a normal distribution with

$$\mu_U = \frac{N_1 N_2}{2}, \text{ and } \sigma_U = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}}.$$

The standard normal statistic  $z = (U - \mu_U)/\sigma_U$  may be used to test the null hypothesis using probabilities from tables for the normal distribution. Several other tests for testing differences in location exist for a variety of research designs, such as the Wilcoxon signed rank test for paired samples, and the Kruskal–Wallis test for comparing differences in location between  $k$  ( $k \geq 3$ ) independent samples.

### Determining Degree of Association

Suppose, two numerical variables  $X$  and  $Y$  have been sampled from the same population and one wishes to know the degree to which they are linearly related. An extremely well-known statistic for this purpose is Spearman's product–moment correlation, denoted  $\rho$  in the population and  $r$  in the sample. Assuming bivariate normality for  $X$  and  $Y$  and a sample of size  $N$ , the null hypothesis that  $\rho = 0$  can be tested using a  $t$  statistic, which can be shown to depend only on  $r$  and  $N$ .

Again, what if variables  $X$  and  $Y$  are ordinal or one is reluctant to assume bivariate normality? An alternative to the product–moment correlation is Spearman's rank correlation  $r_s$ . Here, the scores on both variables are replaced by the corresponding ranks, which for brevity we also denote  $X$  and  $Y$ . For example, for a respondent indexed  $i$ , the pair of ranks ( $X_i = 7, Y_i = 15$ ) show that (s)he was ranked 7 on variable  $X$  and 15 on variable  $Y$ . The difference between paired ranks equals  $d_i = 7 - 15$ , and so on for other respondents. Tied scores are replaced by the average of the ranks assigned, had no ties been observed. The product–moment correlation between the two rankings  $X$  and  $Y$  provide the values of  $r_s$ . Ignoring ties for simplicity, and using differences  $d$  between paired ranks,  $r_s$  equals

$$r_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)}.$$

For small samples, the null hypothesis that  $\rho_s = 0$  can be evaluated by considering all possible rank orderings or permutations of  $Y$  given a particular rank ordering of  $X$ , and determining the value of  $r_s$  for each pair of ranks  $X$  and  $Y$ . As under the null hypothesis, each permutation of  $Y$  has the same probability,  $(N!)^{-1}$ , the probability of a particular value for  $r_s$  simply is a multiple of  $(N!)^{-1}$ , and a table based on the distribution of  $r_s$  given  $N$  is readily prepared. This table can be used for hypothesis testing. For, say,  $N \leq 30$ , this is feasible but for larger  $N$  one runs into combinatorial problems. For  $N \geq 10$  (Hays, 1981: 598), a  $t$ -test may be used for testing the null hypothesis. Other coefficients exist for expressing degree of association, such as Kendall's  $\tau_b$ , which corrects for tied observations (Liebetrau, 1983: 51–53).

When  $X$  and  $Y$  are nominal, a two-way contingency table is set up with cells for all combinations  $(x, y)$  and

frequency counts in each cell, which reflect how often a particular combination is observed in the sample. The null hypothesis of no association may be tested using a chi-squared statistic. The strength of the association may be expressed by several coefficients, for example, the phi-coefficient when both  $X$  and  $Y$  have two categories, and Cramér's  $V$  when the number of categories is greater and not necessarily equal for  $X$  and  $Y$  (Liebetrau, 1983).

### Bootstrap and Permutation Tests

When the population distribution of the observations is unknown, it may happen that the sampling distribution of a statistic of interest, say  $T$ , derived from the data also is unknown. In this case, a computer-intensive way of approximating the sampling distribution of  $T$  is to draw a large number of samples of size  $N$  with replacement from the sample, and then compute statistic  $T$  for each so-called bootstrap sample. The distribution of  $T$  across the bootstrap samples can be used to determine a confidence interval for parameter  $\tau$  of which  $T$  is the estimate, for example, by identifying the 2.5th and the 97.5th percentile and using these as lower and upper bounds, respectively. This procedure is known as the nonparametric bootstrap (Efron and Tibshirani, 1993). Statistic  $T$  can be any quantity of interest, such as the median or the range of a distribution, a difference between group medians, or an association measure between scores.

An example of a related procedure is known as the permutation test (also, see the discussion of the test of the null hypothesis that Spearman's  $\rho_s = 0$ ). In general, for null hypothesis  $\tau = 0$ , the statistic  $T$  is determined for all possible arrangements of the sampled data, which are also known as permutations. The distribution of  $T$  across these permutations is the sampling distribution of  $T$  under the null hypothesis. If one wishes testing the null hypothesis  $\tau = 0$  at the 5% level against the alternative that  $\tau > 0$ , then the 95th percentile of the sampling distribution defines  $T_{\text{crit}}$ . If the observed value of  $T$  in the original sample, denoted  $T_{\text{obs}}$ , exceeds  $T_{\text{crit}}$ , the null hypothesis is rejected. The number of possible permutations can become excessively large, even for high-powered computers, and then a large random sample of permutations may be used instead. For some statistics  $T$ , the distribution of  $T$  approaches a known distribution in larger samples, as with Spearman's  $r_s$ , which approaches a Student's  $t$ -distribution.

### Specific Nonparametric Methods: Nonparametric Regression

As regression analysis is regularly used in educational research, this section presents nonparametric regression analyses as an illustration of a more advanced nonparametric



statistical method. First, the parametric multiple regression analysis is discussed. This method relates a response variable, denoted  $Y$ , linearly to one or more explanatory variables, denoted  $X_j (j = 1, \dots, m)$ , such that

$$Y = \alpha + \sum \beta_j X_j + \varepsilon.$$

In the regression equation, the regression intercept is denoted by  $\alpha$ , the regression coefficients by  $\beta_j (j = 1, \dots, m)$ , and the residual error by  $\varepsilon$ , which is assumed to have 0 mean for fixed values of the explanatory variables. The regression parameters are estimated fitting the regression equation to the data of  $N$  observations (indexed by  $i$ ) so as to minimize  $\sum \varepsilon_i^2$ . Hypotheses of interest are whether  $\beta_j = 0 (j = 1, \dots, m)$  against a one-sided or a two-sided alternative, and whether the amount of variance explained by the model, denoted  $R^2$ , equals 0 (i.e.,  $R^2 = 0$ ), against the alternative that it is positive. For testing these hypotheses, it is assumed (among others) that the conditional distribution of  $Y$  is normal, with mean  $\alpha + \sum \beta_j X_j$  and constant variance  $\sigma^2$ . This amounts to assuming that  $\varepsilon \sim N(0, \sigma^2)$ .

### An Overview of Nonparametric Regression Methods

Suppose, the researcher has insufficient evidence to support linearity, normality, and equal variance, or his/her substantive theory does not imply this degree of structure. Then, (s)he could try nonparametric regression methods to study the relationships between variables. We illustrate that the exact shape as estimated directly from the data may provide interesting information about relationships.

For simplicity, we consider only models with one explanatory variable,  $X$  (see Fox, 1997, 2000b, for many generalizations). Then, when one expects  $Y$  and  $X$  not to be linearly related, the simplest option is to compute the mean  $\bar{Y}$  conditional on separate values of  $X$ , and take the mean as the regression estimate; that is,  $\hat{Y}|x = \bar{Y}|x$  is computed and a graph is drawn that connects adjacent conditional estimates  $\hat{Y}|x$ . Problems may arise when  $X$  is continuous and many different values are observed so that only few values of  $Y$  are tied to one value of  $X$ , when the number of different values of  $X$  is manageable but the total sample size is relatively small, which leads to the same problem, or when the combination of both occurs.

The problem of too few observations per value of  $X$  may be accommodated by the use of so-called bins. Order all observations according to  $X$  from small to large, and define a subset of smallest adjacent  $X$  values to form the first bin, then a subset of the next adjacent values of  $X$  to form the second bin, and so on. Bins are indexed  $k = 1, \dots, K$ . Then,  $\bar{Y}$  is computed based on all observations in the  $k$ th bin ( $k = 1, \dots, K$ ). The resulting values  $\hat{Y}|\text{bin}(k)$  are plotted in a graph, which often appears jagged. Important decisions in binning concern the

width of a bin and the minimum number of observations in a bin. These decisions affect both the bias and the variance of the estimate of mean  $Y$ , and smaller bias often implies greater variance and vice versa.

Bins may also be defined as overlapping windows, where each next window has moved further to the right across the scale of  $X$ , so that values of  $X$  enter the window from the right and exit the window from the left. As for bins, windows may have either a fixed width defined by values of  $X$  or they contain a fixed number of  $n + 1$  observations. In local averaging, for each window  $\bar{Y}$  is computed on the basis of all the observations on  $X$  that are in the window. When the number of observed values on  $X$  is small, the resulting graph is jagged, as with binning. Windows that contain a fixed number of  $n + 1$  observations can be moved across all  $N$  observations on  $X$  after these observations have been ordered from small to large, such that each observation can be the central focal value  $x_0$  of a window once. The other observations in the window can be defined to be the  $n$  nearest neighbors of  $x_0$ . Then,  $N$  conditional estimates  $\hat{Y}|x_0$  (the conditioning is on  $x_0$  plus its  $n$  nearest neighbors) lead to a relatively smooth curve. As with binning, either the width of the window or the size of  $n$  must be determined. Another problem is that with fixed  $n$ ,  $x$  values near the endpoints have more-similar neighborhoods than other  $x$  values, so that the regression curve tends to flatten near the endpoints.

Kernel smoothing takes local averaging a step further by differentially weighing the  $y$  values corresponding to neighbors of focal point  $x_0$ , such that  $y$  values of neighbors close to  $x_0$  receive more weight than  $y$  values of neighbors further away in the window. Denoting weights by  $w_i$  the local average is obtained by means of

$$\hat{Y}|x_0 = \frac{\sum_{i=1}^N w_i y_i}{\sum_{i=1}^N w_i}.$$

Weight  $w_i$  is the kernel function, defined as  $w_i = K[(x_i - x_0)/b]$  ( $b$  is explained shortly). Several choices are possible but a convenient choice is the Gaussian kernel,

$$w_i = K\left(\frac{x_i - x_0}{b}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - x_0/b)^2},$$

in which  $b$  is called the bandwidth. Small bandwidth values produce jittery curves showing too much random detail (too much variance) and large values produce smooth curves that erase salient features of the regression (too much bias). Thus, finding a value of  $b$  that counterbalances bias and precision is an important topic.

The data within a window may be distributed approximately equally dense but they may also pile up on one side of the window. In the latter case, estimate  $\hat{Y}|x_0$  may be heavily biased. Local polynomial regression provides a

better estimate by fitting a polynomial regression model to the data in the window,

$$Y = \alpha + \beta_1(X - x_0) + \beta_2(X - x_0)^2 + \cdots + \beta_p(X - x_0)^p + \varepsilon,$$

using weighted least squares, such that  $\sum_{i=1}^n w_i^2 \varepsilon_i^2$  is minimized so as to obtain estimated regression parameters. Then, for this window  $\hat{Y}|x_0 = \hat{\alpha}$  is the point on the regression curve, and an estimated curve is obtained as the window moves along. As for higher  $p$ , multicollinearity sets in,  $p = 1, 2$ , and  $3$  are convenient practical choices.

With spline regression (Marsh and Cormier, 2002; Ramsay and Silverman, 1997), we return to binning but instead of taking an unweighted average we now fit a polynomial regression model,

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \cdots + \beta_p X^p + \varepsilon,$$

to the data in the bin such that the regression curves in adjacent bins connect smoothly to constitute one smooth curve across the bins. Not only should curves from adjacent bins connect the points on the boundaries of the bins, also called knots, but in passing a knot, the slope of the curve should not change abruptly, meaning that the first derivative is smooth at the knot, and the curvature of the curve, for example, should not change abruptly from smooth to rough, meaning that the second derivative also is smooth at the knot. This requires a polynomial to have at least degree  $p = 3$ , because linear spline functions ( $p = 1$ ) only connect straight lines between bins, and quadratic spline functions ( $p = 2$ ) in addition only guarantee a smooth change of the slope but do not prevent sudden breaks in this change. See Green and Silverman (1994) for methods that control the balance between bias and variance and prevent curves from becoming jumpy.

Both local polynomial regression and spline regression are extremely flexible and overcome many of the weaknesses of other nonparametric regression methods, such as the flattening of curves near the endpoints. Both methods use parametric functions to adequately describe interesting features of relationships but not to hypothesize that relationships are naturally linear, quadratic, and so on. Inference with respect to nonparametric regression is based on estimated confidence bands around the regression curve. Several possibilities exist, the nonparametric bootstrap being one of them. In addition, nested models may be tested against one another.

### An Application of Kernel Smoothing to Educational Test Data

Nonparametric regression methods find their application, for example, in a large and important class of models for educational measurement, known as item response models, discussed elsewhere in this encyclopedia. These models use the scores (binary, nominal, and

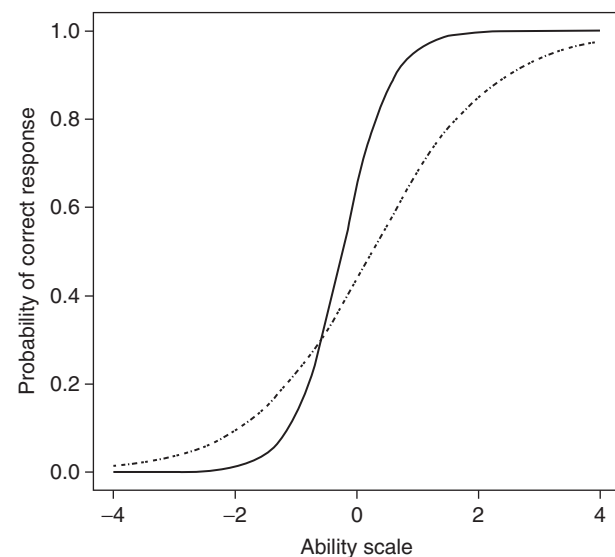
ordered) of a large sample of students (often,  $N \gg 500$ ) on several items that measure the same ability, for constructing an ability scale on which students can be located. Each item may present, for example, an arithmetic problem to the student and the scores on each of the items may be driven by a student's ability to perform well on such problems.

The key feature of item response models is the nonlinear regression of each separate item score on the ability. Let the random variable for an item score be denoted  $X_j$ . It is further assumed that the ability is represented by a latent variable, denoted by  $\theta$ , the scores on which are inferred from the student's observed item scores through the estimation of the item response model. For  $X_j = 0, 1$  (e.g., incorrect/correct solutions), a typical example of a regression function is the 2-parameter logistic function (e.g., Van der Linden and Hambleton, 1997),

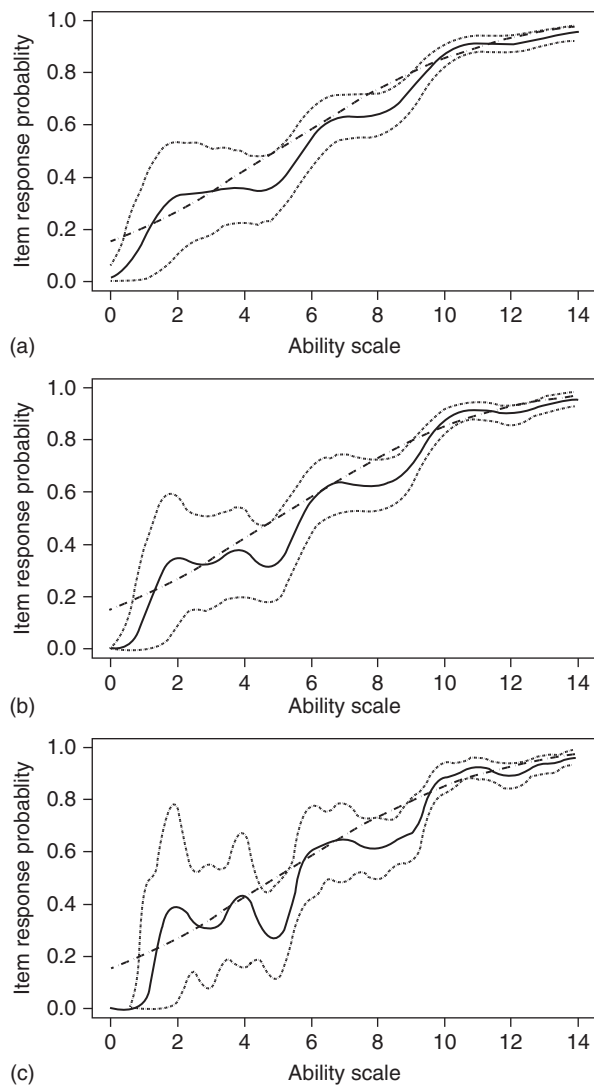
$$P(X_j = 1|\theta) = \frac{e^{a_j(\theta - b_j)}}{1 + e^{a_j(\theta - b_j)}};$$

see Figure 2 for the graphical representation of two logistic functions. Parameter  $b_j$  gives the value of  $\theta$  for which the probability of a correct solution equals 0.5, and parameter  $a_j$  is monotonically related to the function's steepest slope, which is in the inflexion point  $(b_j, 0.5)$ . For parameter estimation, marginal maximum likelihood (MML) may be used, which assumes that  $\theta$  is normal when the item parameters are estimated, and then assumes that the estimated item parameters are fixed when for each individual an ability score  $\theta$  is estimated.

An important question is whether a 2-parameter logistic regression function and a normal  $\theta$  are not unduly restrictive for the problem of interest (Junker and Sijtsma, 2001).



**Figure 2** Two 2-parameter logistic functions: solid curve has parameters  $b_j = -0.25$  and  $a_j = 2$ , and dashed curve has parameters  $b_j = 0.25$  and  $a_j = 1.0$ .



**Figure 3** Estimated item response functions (solid curves) with 95% confidence envelopes (dotted curves): bandwidth parameters (a)  $h = 2$ ; (b)  $h = 1.5$ , and (c)  $h = 1.0$ . The dashed curve in each figure is the estimated 2-parameter logistic function.

Indeed, in the context of item response theory, binning, kernel smoothing, and spline regression each have been proposed and used as alternatives for logistic and other parametric functions. **Figure 3** shows the estimated kernel regression functions of an item response function from a 15-item arithmetic test, for three levels of the bandwidth parameter:  $b = 2$  (**Figure 3(a)**),  $b = 1.5$  (**Figure 3(b)**), and  $b = 1$  (**Figure 3(c)**). The dashed curve is a 2-parameter logistic curve based on MML estimates of the two item parameters,  $a_j$  and  $b_j$ . The solid curve is the estimate based on kernel smoothing, and the dotted curves are the 90% confidence envelopes. As  $b$  becomes smaller, the regression curve becomes jumpier. Each nonparametric estimate suggests that the logistic estimate often seems

to overestimate the true-response probabilities, and does not adequately capture the slope of the regression function at different ranges of the ability scale.

The results illustrate the usefulness of nonparametric data analysis for revealing systematic trends in the data that might have gone unnoticed had parametric methods been used. For example, the nonparametric regression estimate shows that the item does not discriminate well in the ability range from score 2 to score 5 whereas the slope parameter of the parametric regression function cannot reveal this kind of detail. Thus, the nonparametric estimate provided a more realistic impression of how the item functions for measuring ability differences between students.

### Further Reading

Introductory textbooks on nonparametric statistics are, for example, [Hollander and Wolfe \(1999\)](#), [Sheskin \(2007\)](#), and [Siegel \(1956; Siegel and Castellan, 1988\)](#). An example of a multivariate nonparametric method that may be of special interest for educational researchers is the nonparametric approach to multiple analysis of variance (MANOVA) ([Puri and Sen, 1971](#)) to test group differences on multiple outcomes (e.g., several cognitive skills). [Finch \(2005\)](#) showed that the nonparametric MANOVA has good statistical properties in situations in which, for example, the normality assumptions underlying the parametric MANOVA are violated. Another example includes the work by [Linting et al. \(2007\)](#), who discuss nonlinear principal components analysis based on the monotone transformation of ordinal variables such that the relationship with other variables is optimized. For nonparametric regression techniques, the availability of high-speed computers has led to a rapid development of computer-intensive methods, such as bootstrapping and resampling, linear and nonlinear smoothing, and graphical methods, which can handle complex multivariate data (e.g., [Akritas and Politis, 2003](#); [Wasserman, 2006](#)).

### Bibliography

- Akritis, M. C. and Politis, D. N. (2003). *Recent Advances and Trends in Nonparametric Statistics*. Amsterdam: Elsevier.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology* **1**, 27–38.
- Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.
- Fox, J. (2000a). *Nonparametric Simple Regression. Smoothing Scatterplots*. Thousand Oaks, CA: Sage.
- Fox, J. (2000b). *Multiple and Generalized Nonparametric Regression*. Thousand Oaks, CA: Sage.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.



- Hays, W. L. (1981). *Statistics*, 3rd edn. New York: Holt, Rinehart, and Winston.
- Hays, W. L. (1994). *Statistics*, 5th edn. Fort Worth, TX: Harcourt Brace College Publishers.
- Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, 2nd edn. New York: Wiley.
- Junker, B. W. and Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement* **25**, 211–220.
- Liebetrau, A. M. (1983). *Measures of Association*. Beverly Hills, CA: Sage.
- Linting, M., Meulman, J. J., Groenen, P. J. F., and Van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods* **12**, 336–358.
- Marsh, L. C. and Cormier, D. R. (2002). *Spline Regression Models*. Thousand Oaks, CA: Sage.
- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.
- Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York: Springer.
- Sheskin, D. J. (2007). *Parametric and Nonparametric Statistical Procedures*, 4th edn. Boca Raton, FL: Chapman and Hall/CRC Press.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. Tokyo: McGraw-Hill Kogakusha.
- Siegel, S. and Castellan, N. J., Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Van der Linden, W. J. and Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In van der Linden, W. J. and Hambleton, R. K. (eds.) *Handbook of Modern Item Response Theory*, pp 1–28. New York: Springer.
- Wasserman, L. A. (2004). *All of Statistics. A Concise Course in Statistical Inference*. New York: Springer.
- Wasserman, L. A. (2006). *All of Nonparametric Statistics*. New York: Springer.