

Psychometric software

Psychometric software is software that is used for psychometric analysis of data from tests, questionnaires, or inventories reflecting latent psychoeducational variables. While some psychometric analyses can be performed with standard statistical software like SPSS, most analyses require specialized tools.

Contents

Sources

Classical test theory

- CITAS
- jMetrik
- Iteman
- Lertap
- TAP
- ViSta-CITA
- psych

Item response theory calibration

- BILOG-MG
- dexter
- Facets
- flexMIRT
- irtoys
- ICL
- jMetrik
- Lertap5
- MULTILOG
- BMIRT
- PARSCALE
- PARAM-3PL
- TESTFact
- WINMIRA 2001
- Winsteps
- Xcalibre
- IATA
- mirt
- Itm
- TAM
- ACER ConQuest

Additional item response theory software

- LinkMIRT
- METRICS
- SimuMIRT
- SimuMCAT
- IRTEQ
- ResidPlots-2

WinGen
ST
POLYST
STUIRT
plink

Decision consistency

Iteman
jMetrik
Lertap

General statistical analysis software

R
SAS
SPSS
S-Plus
Stata

See also**References**

Sources

There exist many free tools developed by researchers and educators. Important websites for free psychometric software include:

- CASMA at the University of Iowa, USA
- REMP at the University of Massachusetts, USA
- Software from Harold Doran
- Software from Brad Hanson
- Software from John Uebersax
- Software from J. Patrick Meyer
- Software directory at the Institute for Objective Measurement
- Software from Lihua Yao
- Software from Larry Nelson
- CRAN Task View: Psychometric Models and Methods

Classical test theory

Classical test theory is an approach to psychometric analysis that has weaker assumptions than item response theory and is more applicable to smaller sample sizes.

CITAS

CITAS (Classical Item and Test Analysis Spreadsheet) is a free Excel workbook designed to provide scoring and statistical analysis of classroom tests. Item responses (ABCD) and keys are typed or pasted into the workbook, and the output automatically populates; unlike some other programs, CITAS does not require any "running" or experience in psychometric analysis, making it accessible to school teachers and professors.

jMetrik

jMetrik is free and open source software for conducting a comprehensive psychometric analysis. It was developed by J. Patrick Meyer at the University of Virginia. Current methods include classical item analysis, differential item functioning (DIF) analysis, confirmatory factor analysis, item response theory, IRT equating, and nonparametric item response theory. The item analysis includes proportion, point biserial, and biserial statistics for all response options. Reliability coefficients include Cronbach's alpha, Guttman's lambda, the Feldt-Gilmer Coefficient, the Feldt-Brennan coefficient, decision consistency indices, the conditional standard error of measurement, and reliability if item deleted. The DIF analysis is based on nonparametric item characteristic curves and the Mantel-Haenszel procedure. DIF effect sizes and ETS DIF classifications are included in the output. Confirmatory factor analysis is limited to the common factor model for congeneric, tau-equivalent, and parallel measures. Fit statistics are reported along with factor loadings and error variances. IRT methods include the Rasch, partial credit, and rating scale models. IRT equating methods include mean/mean, mean/sigma, Haebara, and Stocking-Lord procedures.

jMetrik also include basic descriptive statistics and a graphics facility that produces bar charts, pie chart, histograms, kernel density estimates, and line plots.

jMetrik is a pure Java application that runs on 32-bit and 64-bit versions of Windows, Mac, and Linux operating systems. jMetrik requires Java 1.6 on the host computer.

Iteman

Iteman is a commercial program specifically designed for classical test analysis, producing rich text (RTF) reports with graphics, narratives, and embedded tables. It calculates the proportion and point biserial of each item, as well as high/low subgroup proportions, and detailed graphics of item performance. It also calculates typical descriptive statistics, including the mean, standard deviation, reliability, and standard error of measurement, for each domain and the overall tests. It is only available from Assessment Systems Corporation.

Lertap

Lertap5 (the 5th version of the Laboratory of Educational Research Test Analysis Program) is a comprehensive software package for classical test analysis developed for use on Windows and Macintosh computers with Microsoft Excel. It includes test, item, and option statistics, classification consistency and mastery test analysis, procedures for cheating detection, and extensive graphics (e.g., trace lines for item options, conditional standard errors of measurement, scree plots, boxplots of group differences, histograms, scatterplots).

DIF, differential item functioning, is supported in the Excel 2010, Excel 2013, Excel 2016, and Excel 2019 versions of Lertap5. Mantel-Haenszel methods are used; graphs of results are provided, including empirical DIF plots.

Users of IRT, item response theory, may make use of four special options: one will produce the data and item control files required by Xcalibre; another sets up an ASCII file for Bilog MG; a third prepares data for processing with SAS, and will write suitable lines of SAS code for use with SAS IRT modules; a fourth interfaces with the Excel version of an R package called "RIRT", allowing users to calibrate items without leaving Excel.

Several sample datasets for use with Lertap and/or other item and test analysis programs are available; these involve both cognitive tests, and affective (or rating) scales. Technical papers related to the application of Lertap5 are also available.

Lertap5 was developed by Larry Nelson at Curtin University and is available from Lertap5.com (<http://www.lertap5.com>).

TAP

TAP (the Test Analysis Program) is a free program for basic classical analysis developed by [Gordon Brooks \(https://people.ohio.edu/brooksg/\)](https://people.ohio.edu/brooksg/) at Ohio University.

ViSta-CITA

ViSta-CITA (Classical Item and Test Analysis) is a module included in the Visual Statistics System (ViSta) that focuses on graphical-oriented methods applied to psychometric analysis. It was developed by Ruben Ledesma, J. Gabriel Molina, Pedro M. Valero-Mora, and Forrest W. Young.

psych

R package. A number of routines for personality, psychometrics and experimental psychology. Functions are primarily for scale construction using factor analysis, cluster analysis and reliability analysis, although others provide basic descriptive statistics. Item Response Theory is done using factor analysis of tetrachoric and polychoric correlations. Functions for simulating particular item and test structures are included. Several functions serve as a useful front end for structural equation modeling. Graphical displays of path diagrams, factor analysis and structural equation models are created using basic graphics. Some of the functions are written to support a book on psychometrics as well as publications in personality research. For more information, see the personality-project.org/r webpage.

Item response theory calibration

Item response theory (IRT) is a psychometric approach which assumes that the probability of a certain response is a direct function of an underlying trait or traits. Various functions have been proposed to model this relationship, and the different calibration packages reflect this. Several software packages have been developed for additional analysis such as equating; they are listed in the next section.

BILOG-MG

BILOG-MG is a software program for IRT analysis of dichotomous (correct/incorrect) data, including fit and differential item functioning. It is commercial, and only available from Scientific Software International or Assessment Systems Corporation.

dexter

dexter (<https://cran.r-project.org/web/packages/dexter/vignettes/dexter.html>), first published February 2017, is an R package intended as a robust and fairly comprehensive system for managing and analyzing test data organized in booklets. The package includes facilities for importing and managing test data, assessing and improving the quality of data through basic test-and-item analysis, fitting an IRT model, and computing various estimates of ability. Many psychometric methods not found elsewhere are provided, such as Haberman's (2007) interaction model generalized for polytomous items, efficient generation of plausible values or scores, new methods for exploratory and confirmatory DIF analysis, support for the 3DC method of standard setting, and many more. The central IRT model is a polytomous generalization of the extended marginal Rasch model. Estimation is by CML or Bayesian techniques. There are two companion packages: dextergui (<https://cran.rstudio.com/web/packages/dextergui/vignettes/dextergui.html>), first published June 2018, and providing an easy graphical interface to the most widely used functions in dexter; and dexterMST (https://cran.r-project.org/web/packages/dexterMST/vignettes/multistage_fundamentals.html), first published July 2018, for managing and analyzing data from multi-stage test designs. All packages are extensively documented both for the beginner as for the professional (see also the blog (<https://dexterities.netlify.com/>)).

Facets

Facets is a software program for Rasch analysis of rater- or judge-intermediated data, such as essay grades, diving competitions, satisfaction surveys and quality-of-life data. Other applications include rank-order data, binomial trials and Poisson counts.

flexMIRT

flexMIRT IRT software is a multilevel, multiple group software package for item analysis, item calibration, and test scoring. The flexMIRT IRT software package fits a variety of unidimensional and multidimensional item response theory models (also known as item factor analysis models) to single-level and multilevel data in any number of groups.

irtoys

irtoys (<https://cran.r-project.org/web/packages/irtoys/vignettes/irtoys-vignette.html>) is an R package first published in 2007 and supporting almost everything in the book but limited to one booklet of dichotomous items. It is good for teaching, smaller projects, as a psychometrician's Swiss knife and as a source of building stones for other projects. The simple syntax files for ICL and BILOG-MG it writes can be studied and modified to handle more complicated problems.

ICL

ICL (IRT Command Language) performs IRT calibrations, including the 1, 2, and 3 parameter logistic models as well as the partial credit model and generalized partial credit model. It can also generate response data. As the name implies, it is completely command code driven, with no graphical user interface.

jMetrik

jMetrik (<https://itemanalysis.com/jmetrik-download/>) is free and open source software for conducting a comprehensive psychometric analysis. It was developed by J. Patrick Meyer at the University of Virginia. Current methods include classical item analysis, differential item functioning (DIF) analysis, item response theory, IRT equating, and nonparametric item response theory. The item analysis includes proportion, point biserial, and biserial statistics for all response options. Reliability coefficients include Cronbach's alpha, Guttman's lambda, the Feldt-Gilmer Coefficient, the Feldt-Brennan coefficient, decision consistency indices, the conditional standard error of measurement, and reliability if item deleted. The DIF analysis is based on nonparametric item characteristic curves and the Mantel-Haenszel procedure. DIF effect sizes and ETS DIF classifications are included in the output. IRT methods include the Rasch, partial credit, and rating scale models estimated via JMLE. jMetrik also provides the 3PL, 4PL, and generalized partial credit models estimated via MMLE. Person scoring methods include MLE, MAP, and EAP. IRT equating methods include mean/mean, mean/sigma, Haebara, and Stocking-Lord procedures.

jMetrik also include basic descriptive statistics and a graphics facility that produces bar charts, pie chart, histograms, kernel density estimates, and line plots.

jMetrik is a pure Java application that runs on 32-bit and 64-bit versions of Windows, Mac, and Linux operating systems. jMetrik requires Java 1.6 on the host computer.

Lertap5

While basically a classical test theory program, Lertap5 (<http://www.lertap5.com>) has support for users of Xcalibre, Bilog-MG, the IRT routines in SAS, and "EIRT", the Excel equivalent of the "RIRT" package. Lertap5 runs an Excel "app", as does EIRT. These two programs combined make item calibration very feasible using just Excel.

MULTILOG

MULTILOG is an extension of BILOG to data with polytomous (multiple) responses. It is commercial, and only available from Scientific Software International or Assessment Systems Corporation.

BMIRT

BMIRT is a free Java multi-purpose application program that conducts item calibrations and ability estimation in a multidimensional, multi-group item response theory (IRT) model framework; it can fit dichotomous or polytomous models, along with mixed models. It supports both exploratory and confirmatory and for both compensatory and noncompensatory MIRT models.

PARSCALE

PARSCALE is a program designed specifically for polytomous IRT analysis. It is commercial, and only available from Scientific Software International or Assessment Systems Corporation.

PARAM-3PL

PARAM-3PL is a free program for the calibration of the 3-parameter logistic IRT model. It was developed by Lawrence Rudner at the Education Resources Information Center (ERIC). The latest release was version 0.93 in August 2012.

TESTFact

Testfact features - Marginal maximum likelihood (MML) exploratory factor analysis and classical item analysis of binary data - Computes tetrachoric correlations, principal factor solution, classical item descriptive statistics, fractile tables and plots - Handles up to 10 factors using numerical quadrature: up to 5 for non-adaptive and up to 10 for adaptive quadrature - Handles up to 15 factors using Monte Carlo integration techniques - Varimax (orthogonal) and PROMAX (oblique) rotation of factor loadings - Handles an important form of confirmatory factor analysis known as "bifactor" analysis: Factor pattern consists of one main factor plus group factors - Simulation of responses to items based on user specified parameters - Correction for guessing and not-reached items - Allows imposition of constraints on item parameter estimates - Handles omitted and not-presented items - Detailed online HELP documentation includes syntax and annotated examples.

WINMIRA 2001

WINMIRA 2001 is a program for analyses with the Rasch model for dichotomous and polytomous ordinal responses, with the latent class analysis, and with the Mixture Distribution Rasch model for dichotomous and polytomous item responses.^{[1][2]} The software provides conditional maximum likelihood (CML) estimation of item parameters, as well as MLE and WLE estimates of person parameters, and person- and item-fit statistics as well as information criteria (AIC, BIC, CAIC) for model selection. The software also performs a parametric bootstrap procedure for the selection of the number of mixture components. A free student version is available from Matthias von Davier's webpage and a commercial version is available.

Winsteps

Winsteps is a program designed for analysis with the Rasch model, a one-parameter item response theory model which differs from the 1PL model in that each individual in the person sample is parameterized for item estimation and it is prescriptive and criterion-referenced, rather than descriptive and norm-referenced in nature.^[3] It is commercially available from Winsteps, Inc. A previous DOS-based version, BIGSTEPS, is also available.

Xcalibre

XCalibre is a commercial program that performs marginal maximum likelihood estimation of both dichotomous (1PL-Rasch, 2PL, 3PL) and all major polytomous IRT models. The interface is point-and-click; no command code required. Its output includes both spreadsheets and a detailed, narrated report document with embedded tables and figures, which can be printed and delivered to subject matter experts for item review. It is only available from Assessment Systems Corporation.

IATA

IATA is a software package for analysing psychometric and educational assessment data. The interface is point-and-click, and all functionality is delivered through wizard-style interfaces that are based on different workflows or analysis goals, such as pilot testing or equating. IATA reads and writes csv, Excel and SPSS file formats, and produces exportable graphics for all statistical analyses. Each analysis also includes heuristics suggesting appropriate interpretations of the numerical results. IATA performs factor analysis, (1PL-Rasch, 2PL, 3PL) scaling and calibration, differential item functioning (DIF) analysis, (basic) computer aided test development, equating, IRT-based standard setting, score conditioning, and plausible value generation. It is available for free from Polymetrika International.

mirt

R package. Analysis of dichotomous and polytomous response data using unidimensional and multidimensional latent trait models under the Item Response Theory paradigm. Exploratory and confirmatory models can be estimated with quadrature (EM) or stochastic (MHRM) methods. Confirmatory bi-factor and two-tier analyses are available for modeling item testlets. Multiple group analysis and mixed effects designs also are available for detecting differential item functioning and modelling item and person covariates.

ltm

R package. Analysis of multivariate dichotomous and polytomous data using latent trait models under the Item Response Theory approach. It includes the Rasch, the Two-Parameter Logistic, the Birnbaum's Three-Parameter, the Graded Response, and the Generalized Partial Credit Models.

TAM

R package. The package includes marginal and joint maximum likelihood estimation of uni- and multidimensional item response models (Rasch, 2PL, Generalized Partial Credit, Rating Scale, Multi Facets), fit statistics, standard error estimation, as well as plausible value imputation and weighted likelihood estimation of ability.

ACER ConQuest

ACER ConQuest is a computer program for fitting both unidimensional and multidimensional item response and latent regression models. It provides data analysis based on a comprehensive and flexible range of item response models (IRM), allowing examination of the properties of performance assessments, traditional assessments and rating scales. ACER ConQuest 4 also offers wider measurement and research community analysis procedures based on the most up-to-date psychometric methods of multifaceted item response models, multidimensional item response models, latent regression models and drawing plausible values.

Additional item response theory software

Because of the complexity of IRT, there exist few software packages capable of calibration. However, many software programs exist for specific ancillary IRT analyses such as equating and scaling. Examples of such software follow.

LinkMIRT

LinkMIRT is a free Java application program that links two sets of item parameters in a multidimensional IRT (MIRT) framework. The software can implement the Stocking and Lord method, the mean/mean method, and the mean/sigma method. Linking by comment-person and by random equivalent-groups design are supported.

METRICS

METRICS (<https://shiny.airast.org/METRICS/>) (Methods for Equating, Testing, Regression, Item Response Theory, and Scoring) is a free web-based application for psychometricians. It was developed by Harold Doran at the American Institutes for Research. The software includes an environment to use interactive visual displays of data along with various data exploratory tools, computation of reliability using Cronbach's alpha, stratified alpha, and Feldt-Raju along with jackknife variances, classical item analysis of test items, IRT-based fit statistics including item fit plots, Yen's Q1 and Q3 statistics, six different equating methods along with visual displays of item characteristic curves after linking and test characteristics curves, a complete IRT-based test scoring engine for scoring unidimensional and multidimensional models given a set of item parameters, classification consistency and classification accuracy methods, linear regression models including fixed effects and mixed effects linear models designed to account for measurement error in the predictor variables, and direct estimation procedures as used in NAEP-style analyses.

SimuMIRT

SimuMIRT is a program that simulates multidimensional data (examinee ability and item responses) for a fixed form (i.e., paper and pencil) test, from a user-specified set of parameters. The rater-effect model is supported.

SimuMCAT

SimuMCAT is a free Java application program that simulates a multidimensional computer adaptive test (MCAT). The user can select from five different MCAT item selection procedures (Volume, Kullback-Leibler information, Minimize the error variance of the linear combination, Minimum Angle, and Minimize the error variance of the composite score with the optimized weight).^[4] Two exposure control approaches are possible: the traditional Sympton-Hetter approach and a maximum exposure control approach. It is also possible to implement content constraints using the Priority Index method. Different stopping rules are implemented with fixed-length test and varying-length test. The user specifies true examinee ability, item pools, and item selection procedures, and the program outputs selected items with item responses and ability estimates. Bayesian and non-Bayesian methods can be specified by the user. The examinees' ability and item pools can also be created from the program by the user specified distributions.

IRTEQ

IRTEQ is a freeware Windows GUI application that implements IRT scaling and equating developed by Kyung (Chris) T. Han. It implements IRT scaling/equating methods that are widely used with the “Non-Equivalent Groups Anchor Test” design: Mean/Mean, Mean/Sigma, Robust Mean/Sigma, and TCC methods.^{[5][6][7][8][9]} For TCC methods, IRTEQ provides the user with the option to choose various score distributions for incorporation into the loss function. IRTEQ supports various popular unidimensional IRT models: Logistic models for dichotomous responses (with 1, 2, or 3 parameters) and the Generalized Partial Credit Model (GPCM) (including Partial Credit Model (PCM), which is a special case of GPCM) and Graded Response Model (GRM) for polytomous responses. IRTEQ can also equate test scores on the scale of a test to the scale of another test using IRT true score equating.^[10]

ResidPlots-2

ResidPlots-2 is a free program for IRT graphical residual analysis. It was developed by Tie Liang, Kyung (Chris) T. Han, and Ronald K. Hambleton at the University of Massachusetts Amherst.

WinGen

WinGen is a free Windows-based program that generates IRT parameters and item responses. Kyung (Chris) T. Han at the University of Massachusetts Amherst.^[11]

ST

ST conducts item response theory (IRT) scale transformations for dichotomously scored tests.

POLYST

POLYST conducts IRT scale transformations for dichotomously and polytomously scored tests.

STUIRT

STUIRT conducts IRT scale transformations for mixed-format tests (tests that include some multiple choice items and some polytomous items).

plink

R package. This package uses item response theory methods to compute linking constants and conduct chain linking of unidimensional or multidimensional tests for multiple groups under a common item design. The unidimensional methods include the Mean/Mean, Mean/Sigma, Haebara, and Stocking-Lord methods for dichotomous (1PL, 2PL and 3PL) and/or polytomous (graded response, partial credit/generalized partial credit, nominal, and multiple-choice model) items. The multidimensional methods include the least squares method and extensions of the Haebara and Stocking-Lord method using single or multiple dilation parameters for multidimensional extensions of all the unidimensional dichotomous and polytomous item response models. The package also includes functions for importing item and/or ability parameters from common IRT software, conducting IRT true score and observed score equating, and plotting item response curves/surfaces, vector plots, and comparison plots for examining parameter drift.

Decision consistency

Decision consistency methods are applicable to criterion-referenced tests such as licensure exams and academic mastery testing.

Iteman

Iteman provides an index of decision consistency as well as a classical estimate of the conditional standard error of measurement at the cutscore, which is often requested for accreditation of a testing program.

jMetrik

jMetrik is free and open source software for conducting a comprehensive psychometric analysis. Detailed information is listed above. jMetrik includes Huynh's decision consistency estimates if cut-scores are provided in the item analysis.

Lertap

Lertap calculates several statistics related to decision and classification consistency, including Livingston's coefficient, the Brennan-Kane dependability index, kappa, and an estimate of $p(o)$, number of correct classifications as a proportion, derived by using the Peng-Subkoviak adaptation of Huynh's method. More detailed information concerning Lertap is provided above, under 'Classical test theory'.

General statistical analysis software

Software designed for general statistical analysis can often be used for certain types of psychometric analysis. Moreover, code for more advanced types of psychometric analysis is often available.

R

R is a programming environment designed for statistical computing and production of graphics. Basic R functionality can be extended through installing contributed 'packages', and a list of psychometric related packages is maintained on the CRAN website ^[12].

SAS

SAS is a commercially available package for statistical analysis and manipulation of data. It is also command-based.

SPSS

SPSS, originally called the Statistical Package for the Social Sciences, is a commercial general statistical analysis program where the data is presented in a spreadsheet layout and common analyses are menu driven.

S-Plus

S-Plus is a commercial analysis package based on the programming language *S*.

Stata

Stata is a commercial package. Stata's implementation of IRT includes 1, 2 and 3 parameter logistic models, graded response models, partial credit and generalized partial credit models, rating scale models, and a nominal response model for unordered categorical responses. It is driven by a control panel that allows the user to specify the model, examine fit numerically and graphically and investigate differential item functioning from a single interface ^[13].

See also

- [Psychological Testing](#)
- [Automatic Item Generation](#)

References

- Rost, J. (1990). **Rasch models** in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- von Davier, M., & Rost, J. (1995). Polytomous mixed **Rasch models**. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments, and applications* (pp. 371-382). New York: Springer.
- Rasch dichotomous model vs. One-parameter Logistic Model **[1]** (<http://www.rasch.org/rmt/rmt193h.htm>). *Rasch Measurement Transactions* **[2]** (<http://www.rasch.org/rmt/>), 2005, 19:3 p. 1032
- Yao, Lihua (2012-05-17). "Multidimensional CAT Item Selection Methods for Domain Scores and Composite Scores: Theory and Applications". *Psychometrika*. **77** (3): 495–523. doi:[10.1007/s11336-012-9265-5](https://doi.org/10.1007/s11336-012-9265-5) (<https://doi.org/10.1007/s11336-012-9265-5>). ISSN 0033-3123 (<https://www.worldcat.org/issn/0033-3123>). PMID 27519778 (<https://www.ncbi.nlm.nih.gov/pubmed/27519778>).
- Loyd & Hoover, 1980
- Marco, 1977
- Linn, Levine, Hastings, & Wardrop, 1981
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22, 144-149.
- Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31, 457-459.
- [3]** (<https://cran.r-project.org/web/views/Psychometrics.html>) CRAN Task View: Psychometric Models and Methods
- [4]** (<https://www.stata.com/manuals/irt.pdf>) Stata's IRT manual online

Retrieved from "https://en.wikipedia.org/w/index.php?title=Psychometric_software&oldid=908687282"

This page was last edited on 31 July 2019, at 09:47 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.