Free Statistical Software

This page contains links to free software packages that you can download and install on your computer for stand-alone (offline, non-Internet) computing. They are listed below, under the following general headings:

- General Packages: support a wide variety of statistical analyses
- Subset Packages: deal with a specific area of analysis, or a limited set of tests
- Curve Fitting and Modeling: to handle complex, nonlinear models and systems
- Biostatistics and Epidemiology: especially useful in the life sciences
- Surveys, Testing and Measurement: especially useful in the business and social sciences
- Excel Spreadsheets and Add-ins: you need a recent version of Excel
- Programming Languages and Subroutine Libraries: customized for statistical calculations; you need to learn the appropriate syntax
- Scripts and Macros: for scriptable packages, like SAS, SPSS, R, etc.
- Miscellaneous: don't fit into any of the other categories
- Other Collections of Links to Free Software

General Packages: No package does everything, but these programs support a wide variety of statistical analyses.

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
OpenStat -- a general stats package for all Windows	SYSTAT 12 -- powerful statistical software ranging

versions (including Win 7 and Win 8) and for Linux systems (under Wine), developed by Bill Miller of Iowa State U, with a **very** broad range of data manipulation and analysis capabilities and an SPSS-like user interface. Bill also has provided an excellent User Manual as an Adobe Acrobat file. For some other interesting programs and documents, check out Bill's Web Site.

SalStat-2 -- a multi-platform, easy-to-use statistical system that provides data management (importing, editing, pivot tables), statistical calculations (descriptive summaries, probability functions, chi-square, t-tests, 1-way ANOVA, regression, correlation, non-parametric tests, Six-Sigma), and graphs (bar, line, scatter, area, histogram, box&whisker, stem, adaptive, ternary scatter, normal probability, quality control).

SOFA (Statistics Open For All) -- an innovative statistics, analysis, and reporting program. Available for Windows, Mac and Linux systems. Has an emphasis on ease of use, learn as you go, and beautiful output. Check out list of features.

ViSta -- a Visual Statistics program for Win3.1, Win 95/NT, Mac and Unix, featuring a Structured Desktop, with features designed to structure and assist the statistical analyst.

PSPP -- a free replacement for SPSS (although at this time it implements only a small fraction of SPSS's

from the most elementary descriptive statistics to very advanced statistical methodology. Novices can work with its friendly and simple menu-dialog; statistically-savvy users can use its intuitive command language. Carry out very comprehensive analysis of univariate and multivariate data based on linear, general linear, and mixed linear models; carry out different types of robust regression analysis when your data are not suitable for conventional multiple regression analysis;compute partial least-squares regression;design experiments, carry out power analysis, do probability calculations on many distributions and fit them to data; perform matrix computations. Provides Time Series, Survival Analysis, Response Surface Optimization, Spatial Statistics, Test Item Analysis, Cluster Analysis, Classification and Regression Trees, Correspondence Analysis, Multidimensional Scaling, Conjoint Analysis, Quality Analysis, Path Analysis, etc. A 30-day evaluation version is available for free download.

Statlets -- a 100% Pure Java statistics program. Should run on any platform (PC, Mac, Unix) that supports Java. The free Academic Version is limited to 100 cases by 10 variables.

WINKS (Windows KWIKSTAT) -- a full-featured, easy-to-use stats package with statistics (means, standard deviations, medians, etc.), histograms, t-tests, correlation, chi-square, regression, nonparametrics,

analyses). But it's free, and will never "expire". It replicates the "look and feel" of SPSS very closely, and even reads native SPSS syntax and files! Some other features...

- Supports over 1 billion cases and over 1 billion variables.
- Choice of terminal or graphical user interface; Choice of text, postscript or html output formats.
- Inter-operates with Gnumeric, OpenOffice.Org and other free software.
- Easy data import from spreadsheets, text files and database sources.
- Fast statistical procedures, even on very large data sets.
- No license fees; no expiration period; no unethical “end user license agreements”.
- Fully indexed user manual.
- Cross platform; Runs on many different computers and many different operating systems.
  Note: For Windows installer, click here.

OpenEpi Version 2.3 -- OpenEpi is a free, web-based, open source, operating-system-independent series of programs for use in public health and medicine, providing a number of epidemiologic and statistical tools. Version 2 (4/25/2007) has a new interface that presents results without using pop-up windows, and has better installation methods so that it can be run without an internet connection. Version 2.2 (2007/11/09) lets users run the

correlation, chi-square, regression, nonparametrics, analysis of variance (ANOVA), probability, QC plots, cpk, graphs, life tables, time series, crosstabs, and more. Works on Windows XP (as well as Windows 2000, NT, 98, ME and 95.) Comes in Basic and Professional editions. Evaluation version available for download.

StudyResult -- (30-day free trial) General statistics package for: paired & unpaired t-test, one-way ANOVA, Fisher's exact , McNemar's, Chi2, Chi^2 homogeneity , life table & survival analysis, Wilcoxon rank-sum & signed-rank, sign test, bioequivalence testing, correlation & regression coefficient tests. Special features for interpreting summary data found in publications (p-values & conf. intervals from summary statistics, converts p-values to CI's & vice versa, what observed results are needed to get a significant result, estimates from publications needed for sample size calculations). Includes equivalence- and non-inferiority testing for most tests.

STATGRAPHICS Plus v5.0 (for Windows) -- over 250 statistical analyses: regression, probit, enhanced logistic, factor effects plots, automatic forecasting, matrix plots, outlier identification, general linear models (random and mixed), multiple regression with automatic Cochrane-Orcutt and Box-Cox procedures, Levene's, Friedman's, Dixon's and Grubb's tests, Durbin-Watson p-values and 1-variable bootstrap estimates, enhanced 3D charts. For *Six Sigma* work:

software in English, French, Spanish, or Italian.

Statext -- Provides a nice assortment of basic statistical tests, with text output (and text-based graphics). Capabilities include: rearrange, transpose, tabulate and count data; random sample; basic descriptives; text-plots for dot, box-and-whiskers, stem-and-leaf, histogram, scatterplot; find z-values, confidence interval for means, t-tests (one and two group, and paired; one- and two-way ANOVA; Pearson, Spearman and Kendall correlation; ;inear regression, Chi-square goodness-of-fit test and independence tests; sign test, Mann-Whitney U and Kruskal-Wallis H tests, probability tables (z, t, Chi-square, F, U); random number generator; Central Limit Theorem, Chi-square distribution.

MicrOsiris -- a comprehensive statistical and data management package for Windows, derived from the OSIRIS IV package developed at the University of Michigan. It was developed for serious survey analysis using moderate to large data sets. **Main features**: handles any size data set; has Excel data entry; imports/exports SPSS, SAS, and Stats datasets; reads ICPSR (OSIRIS) and UNESCO (IDAMS) datasets; data mining techniques for market analysis (SEARCH --very fast for large datasets); interactive decision tree for selecting appropriate tests; database maniuplation (dictionaries, sorting, merging, consistency checking, recoding, transforming) extensive statistics (univariate, staccerplot, cross-tabs, ANOVA/MANOVA, log-linear,

estimates, enhanced 3D charts. For *Six Sigma* work: gage linearity and accuracy analysis, multi-vari charts, life data regression for reliability analysis and accelerated life-testing, long-term and short-term capability assessment estimates. Two free downloads are available: full-function but limited-time(30 days), and unlimited-time but limited-function (no Save, no Print, not all analyses).

NCSS-2007 (Statistical Analysis System), PASS-2008 (Power and Sample Size, and GESS (Gene Expression software for Micro-arrays) for Windows. Free 7-day evaluation versions.

MiniTab -- a powerful, full-featured MS Windows package, with good coverage of industrial / quality control analyses. The free Version 12 Demo expires after 30 days.

InStat (Instant Statistics), a full-featured statistics package from GraphPad Software. Demo version disables printing, saving and exporting capabilities. Demo available for Windows only; commercial version available for Windows and Mac.

Prism -- from GraphPad Software. Performs basic biostatistics, fits curves and creates publication quality scientific graphs in one complete package (Mac and Windows). Windows demo is fully-functional for 30 days, then disables printing, saving and exporting; Mac demo always disables these functions.

correlation/regressionMCA, MNA, binary segmentation, cluster, factor, MINISSA, item analysis, survival analysis, internal consistency); online, web-enabled users manual; requires only 6MB RAM; uses 12MB disk, including manual. Fully-functional version is free; the authors would appreciate a small donation to support ongoing development and distribution.

Gnumeric -- a high-powered spreadsheet with better statistical features than Excel. Has 60 extra functions, basic support for financial derivatives (Black Scholes) and telecommunication engineering, advanced statistical analysis, extensive random number generation, linear and non-linear solvers, implicit intersection, implicit iteration, goal seek, and Monte Carlo simulation tools.

Statist -- a compact, portable program that provides most basic statistical capabilities: data manipulation (recoding, transforming, selecting), descriptive stats (including histograms, box&whisker plots), correlation & regression, and the common significance tests (chi-square, t-test, etc.). Written in C (source available); runs on Unix/Linux, Windows, Mac, among others.

Tanagra -- a free (open-source) data-mining package, which supports the standard "stream diagram" paradigm used by most data-mining systems. Contains components for Data source (tab-delimited text), Visualization (grid, scatterplots), Descriptive statistics (cross-tab, ANOVA, correlation), Instance selection (sampling, stratified),

CoStat 6.2 -- an easy-to-use program for data manipulation and statistical analysis, from CoHort Software. Use a spreadsheet with any number of columns and rows of data: floating point, integer, date, time, degrees, text, etc. Import ASCII, Excel, MatLab, S+, SAS, Genstat, Fortran, and others. Has ANOVA, multiple comparisons of means, correlation, descriptive statistics, analysis of frequency data, miscellaneous tests of hypotheses, nonparametric tests, regression (curve fitting), statistical tables, and utilities. Has an auto-recorder and macro programming language. Callable from the command line, batch files, shell scripts, pipes, and other programs; can be used as the statistics engine for web applications. Free time-limited demo available.

Feature selection and construction, Regression (multiple linear), Factorial analysis (principal components, multiple correspondence), Clustering (kMeans, SOM, LVQ, HAC), Supervised learning (logistic regr., k-NN, multi-layer perceptron, prototype-NN, ID3, discriminant analysis, naive Bayes, radial basis function), Meta-spv learning (instance Spv, arcing, boosting, bagging), Learning assessment (train-test, cross-validation), and Association (Agrawal a-priori). (French-language page here)

Dap -- a statistics and graphics package developed by Susan Bassein for Unix and Linux systems, with commonly-needed data management, analysis, and graphics (univariate statistics, correlations and regression, ANOVA, categorical data analysis, logistic regression, and nonparametric analyses). Provides some of the core functionality of SAS, and is able to read and run many (but not all) SAS program files. Dap is freely distributed under a GNU-style "copyleft".

PAST -- an easy-to-use data analysis package aimed at paleontology including a large selection of common statistical, plotting and modelling functions: a spreadsheet-type data entry form, graphing, curve fitting, significance tests (F, t, permutation t, Chi-squared w. permutation test, Kolmogorov-Smirnov, Mann-Whitney, Shapiro-Wilk, Spearman's Rho and Kendall's Tau tests, correlation, covariance, contingency tables, one-way ANOVA, Kruskal-Wallis test), diversity and similarity indices & profiles, abundance model fitting, multivariate

statistics, time series analysis, geometrical analysis, parsimony analysis (cladistics), and biostratigraphy.

AM -- a free package for analyzing data from complex samples, especially large-scale assessments, as well as non-assessment survey data. Has sophisticated stats, easy drag & drop interface, and integrated help system that explains the statistics as well as how to use the system. Can estimate models via marginal maximum likelihood (MML), which defines a probability distribution over the proficiency scale. Also analyzes "plausible values" used in programs like NAEP. Automatically provides appropriate standard errors for complex samples via Taylor-series approximation, jackknife & other replication techniques.

Instat Plus -- from the University of Reading, in the UK. (Not to be confused with Instat from GraphPad Software.) An interactive statistics package for Windows or DOS.

WinIDAMS -- from UNESCO -- for numerical information processing and statistical analysis. Provides data manipulation and validation facilities classical and advanced statistical techniques, including interactive construction of multidimensional tables, graphical exploration of data (3D scattergram spinning, etc.), time series analysis, and a large number of multivariate techniques.

SSP (Smith's Statistical Package) -- a simple, user-

friendly package for Mac and Windows that can enter/edit/transform/import/export data, calculate basic summaries, prepare charts, evaluate distribution function probabilities, perform simulations, compare means & proportions, do ANOVA's, Chi Square tests, simple & multiple regressions.

Also, check out R and Ox, described in the Programming Languages section below.

Dataplot -- (Unix, Linux, PC-DOS, Windows) for scientific visualization, statistical analysis, and non-linear modeling. Has extensive mathematical and graphical capabilities. Closely integrated with the NIST/SEMATECH Engineering Statistics Handbook.

WebStat -- A Java-based statistical computing environment for the World Wide Web. Needs a browser, but can be downloaded and run offline.

Regress+ -- A professional package (Macintosh only) for univariate mathematical modeling (equations and distributions). The most powerful software of its kind available anywhere, with state-of-the-art functionality and user-friendliness. Too many features to even begin to list here.

SISA -- Simple Interactive Statistical Analysis for PC (DOS) from Daan Uitenbroek. An excellent collection of individual DOS modules for several statistical

calculations, including some analyses not readily available elsewhere.

Statistical Software by Paul W. Mielke Jr. -- a large collection of executable DOS programs (and Fortran source). Includes: Matrix occupancy, exact g-sample empirical coverage test, interactions of exact analyses, spectral decomposition analysis, exact mrbp (randomized block) analyses, exact multi-response permutation procedure, Fisher's Exact for cross-classfication and goodness-of-fit, Fisher's combined p-values (meta analysis), largest part's proportion, Pearson-Zelterman, Greenwood-Moran and Kendall-Sherman goodness-of-fit, runs tests, multivariate Hotelling's test, least-absolute-deviation regression, sequential permutation procedures, LAD regression, principal component analysis, matched pair permutation, r by c contingency tables, r-way contingency tables, and Jonkheere-Terpstra.

IRRISTAT -- for data management and basic statistical analysis of experimental data (Windows). Primarily for analysis of agricultural field trials, but many features can be used for analysis of data from other sources. Includes: Data management with a spreadsheet , Text editor, Analysis of variance, Regression, Genotype x environment interaction analysis, Quantitative trait analysis, Single site analysis, Pattern analysis, Graphics, Utilities for randomization and layout, general factorial EMS, and orthogonal polynomial.

Subset Packages: Each of these programs deals with a specific area of statistics (such as power analysis or mulitvariate analysis), or carries out a specific test or computation.

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
PS -- a **well-implemented** Windows program for **power and sample size calculations** from Vanderbilt Univ Med Ctr. Handles dichotomous, continuous,or survival response measures, which are analyzed by chi-square or Fisher Exact tests, Student t tests, and log-ranks tests, respectively. The alternative hypothesis may be specified either in terms of differing response rates, means, or survival times, or in terms of relative risks or odds ratios. Studies with dichotomous or continuous outcomes may involve either a matched or independent study design. **The latest version also handles Mantel-Haenszel tests.** Can determine sample size for a specified power, power for a specified sample size, or the specific alternative hypotheses that can be detected with a given power and sample size. Produces graphs of relationships between power, sample size and detectable alternative hypotheses (with any two of these variables on x & y, and the third variable generating a family of curves on a single graph). Linear or logarithmic axes may be specified. Can print professional-quality power charts. **The latest version also provides a concise and precise verbal description of each power analysis, which you can copy and paste into the Power/Sample-Size section of your proposal or protocol.**	Data Preparator -- handles the "pre-processing" chores of getting a data file ready for analysis... data cleaning, discretization, numeration, scaling, attribute selection, missing values, outliers, statistics, visualization, balancing, sampling, row selection, and several other tasks, with a user-friendly graphical interface. It's written in Java, and runs on Windows, Mac OS/X, and Linux. The free demo has all features enabled, and will handle up to 200 cases.
G*Power 3 -- a **very general** Power Analysis program for Windows and Macintosh. Performs exact analysis for 6 types of correlation tests, 3 types of bivariate regression tests, 1-group and 2-group comparison of means tests	StatCalc (30-day free trial download) -- a handy desk-top tool and instructional aid that

(parametric and non-parametric), 4 types of multiple regression tests, logistic regression, poisson regression, ordinary and repeated-measures ANOVAs, ANCOVAs, MANOVAs, multivariate T2 and MANOVAs, 8 types of tests of proportions (McNemar, Fisher, etc.), 1-group and 2-group variance tests, and completely generic tests involving the binomial, normal, t, chi-square, and F distributions. Computes power, sample sizes, alpha, beta, and alpha/beta ratios. Has a comprehensive web-based tutorial and reference manual.

Factor -- a comprehensive **factor analysis program**. Provides univariate and multivariate descriptive statistics of input variables (mean, variance, skewness, kurtosis), Var charts for ordinal variables, dispersion matrices (user defined , covariance, pearson correlation, polychoric correlation matrix with optional Ridge estimates). Uses MAP, PA (Parallel Analysis), and PA - MBS (with marginally bootstrapped samples) to determine the number of factors/components to be retained. Performs the following factor and component analyses: PCA, ULS (with Heywood correction), EML, MRFA, Schmid-Leiman second-order solution, and Factor scores. Rotation methods: Quartimax, ,Varimax , Weighted Varimax, Orthomin , Direct Oblimin, Weighted Oblimin, Promax, Promaj , Promin, and Simplimax. Indices used in the analysis: dispersion matrix tests (determinant, Bartlett's, Kaiser-Meyer-Olkin), goodness of fit: Chi-Square ,non-normed fit index, comparative fit index, goodness of fit index, adjusted GFI, RMS error of approx, and estimated non-centrality parameter (NCP), reliabilities of rotated components , simplicity indices: Bentler's, and loading simplicity index. Provides mean, variance and histogram of fitted and standardized residuals, and automatic detection of large standardized residuals.

KEYFINDER -- a menu-driven interactive program for generating, randomizing and tabulating blocked and/or fractional-replicate factorial designs in completely general situations. It can generate blocked and/or fractional-replicate designs with user-specified confounding and aliasing properties. KEYFINDER runs on all

transforms from a standard calculator to a collection of modules that calculate statistics, graph distributions, and provide statistical help with definitions, formulas, and interpretation. (Windows)

WinSPC (30-day free trial)-- statistical process control software to:

- collect quality data from devices, shop-floor machines, data sources, other software systems, or via keyboard;
- monitor plant-wide operations from a single screen, and initiate corrective actions for out-of-control processes (trigger alarm, send email, page an operator, or shut down an out-of-control machine);
- perform statistical analysis to solve problems, optimize processes, and create

versions of Windows. You can download the Version 3.3 Overview document, in PDF format, here. To obtain a free copy of the program and manual, send an e-mail to the author: Peter.Zemroch@shell.com

Weka -- a collection of machine learning algorithms for data mining tasks, implemented in Java. Can be executed from a command-line environment, or from a graphical interface, or can either be called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization, and is well-suited for developing new machine learning schemes.

StatCalc -- a **PC calculator** that computes table values and other statistics for **34 probability distributions**. Also includes some **nonparametric table values**, **tolerance factors**, and **bivariate normal distribution**. A help file is provided for each distribution.

Scientific Calculator - ScienCalc program contains high-performance arithmetic, trigonometric, hyperbolic and transcendental calculation routines. All the function routines therein map directly to Intel 80387 FPU floating point machine instructions.

Distributions -- Windows program allows for the analysis of discrete single dimension distributions. The program is based on various manipulations of the poisson, binomial and hypergeometric distribution. Available are the probability of an observed number of cases given a certain null hypothesis, the calculation of exact poisson, binomial or hypergeometric confidence intervals, the exact and approximate size of a population using catch-recatch methodologies, the full analysis of a Poisson distributed rate ratio, Fieller analysis, and two versions of the negative binomial distribution can be used in various ways. Beside the exact procedures there are also various approximate procedures available. From the

quality reports.

The Unscrambler -- multivariate data analysis software for exploratory statistics, regression analysis, classification, prediction, principal components analysis (PCA), partial least squares regression (PLSR) analysis and three-way PLS regression and experimental design. Free 30-day evaluation copy available.

ADDPLAN -- software for sample size calculation and adaptive monitoring of clinical trials. Handles traditional single fixed sample designs, survival analyses, proportions, means, non-inferiority, flexible (adaptive) designs, group-sequential designs, ?-spending function designs, simultaneous control of Type I & II error, adaptive sample sze recalculation, recursive designs based on conditional Type I error, interim monitoring &

Downloads section of the QuantitativeSkills web site.

Multinomial -- This Windows program is the exact solution to the Chi-square Goodness of fit test of testing for a difference between an observed and an expected distribution in a one-dimensional array. For example, the test can be used to compare the distribution of diseases in a certain locality with an expected distribution on the basis of national or international experiences using an ICD classification. In a two-category array the multinomial test provides a two-sided solution for the Binomial test. For example, Multinomial {10 20 0.20 0.80} gives the two-sided probability (0.105) for the single sided Binomial {0.20 10 30} probability (0.061). The multinomial allows you to work with empty '0' observation cells although you must have an expectation about a cell. From the Downloads section of the QuantitativeSkills web site.

Tables -- a Windows program for the analysis of tables with up to 2*7 and 3*3 cells. The program allows for exact and approximate statistics to be calculated for traditional, ordinal and agreement tables. Fisher exact, Number Needed to Treat, Proportional Reduction in Error Statistics, Normal Approximations, Four different Chi-squares, Gamma, Odds-ratio, t-tests and Kappa are among the many statistical procedures available. From the Downloads section of the QuantitativeSkills web site.

MorePower -- another **well-implemented** power/sample-size calculator for any ANOVA design, for 1- and 2-sample t-tests, and for 1- and 2-sample binomial testing (sign test, chi-square test).

EqPlot -- Equation graph plotter program plots 2D graphs from equations. The application comprises algebraic, trigonometric, hyperbolic and transcendental functions.

error, interim monitoring & analysis, classical stopping boundaries, repeated and end-of-trial conf. inter's and p-values, and a powerful simulation tool. Free 30-day limited-function trial version available for download.

Statistics Problem Solver -- tutoring software that not only solves statistical problems, but also generates step-by-step solutions in order to help students understand how to solve statistical problems. Includes: Histograms, Binomial, Poisson, Exponential, Continuous, Group Comparison and Test of Significance. Two other similar tutorial programs are available from the same company: A&G Grapher -- for plotting any 2D or 3D equation typed into the program. Graphs can be customized in color, scale, resolution, etc., and can be exported or copied-and-pasted into other application. Also calculates slope, area

BlockTreat -- a Java program that implements a very general Monte Carlo procedure that performs non-parametric tests (based on random permutations, not ranks) for block and treatment tests, tests with matching, k-sample tests, and tests for independence between any two random variables. Designs may be incomplete and unbalanced, or even have supernumerary entries. The tests are "exact", in the Monte-Carlo sense -- they can be made as accurate as desired by specifying enough random shuffles.

PCP (Pattern Classification Program) -- a machine-learning program for supervised classification of patterns (vectors of measurements). PCP implements: Fisher's linear discriminant, dimensionality reduction using SVD, PCA, feature subset selection, Bayes error estimation, parametric classifiers (linear and quadratic), L-S (pseudo-inverse) linear discriminant, k-Nearest Neighbor, neural networks (Multi-Layer Perceptron), SVM, model selection for SVM, cross-validation, and bagging (committee) classification. Supports interactive (keyboard-driven menus) and batch processing.

PEPI -- a collection of 43 small DOS / Windows programs that perform a large assortment of statistical tests. They can be downloaded individually, or as a single ZIP file. (A new Windows version is being developed; the test version can be downloaded here.) They were written to accompany the book *Computer Programs for Epidemiologic Analyses: PEPI v. 4.0*, by Abramson and Gahlinger, which is available for purchase. A freely-accessible article describing the new features of WinPEPI can be accessed here. The programs include: p-value adjustments for multiple significance tests; Attributable and Prevented Fractions: Case-Control Studies; Analysis of 2 x 2 Tables; Chi-square Tests of Association; Combining Measures of Association or Probabilities; Confidence Intervals; Aids to Use of Pearson's Correlation Coefficients; ifference Between Rates, Proportions or Means; Direct Standardization; Exact Test for a 2 x K Table; Tests for Goodness of Fit ; Fitting of Poisson and Binomial Distributions; Appraisal of

under the curve, tracing and matrix transformation. Calculus Problem Solver -- differentiates any arbitrary equation and outputs the result, providing detailed step-by-step solutions in a tutorial-like format. Can also initiate an interactive quiz in which you can solve differentiation while the computer corrects your solutions.

Power and Precision -- (30-day free trial download) An "industrial strength" program for calculating power, sample size, and attainable precision for: t-tests and z-tests (1-group, 2-group, paired, unpaired), Proportions (1-group, 2-group Chi Square or Fisher Exact, McNemar, KxC, Sign test), Correlations (1-group and 2-group), ANOVA / ANCOVA (1-, 2-, or 3-way), Multiple hierarchical linear regression (covariates / main effects, interactions, dummy-coded variables, polynomial), R^2

Frequency Distribution ; Indirect Standardization; Agreement Between Categorical Ratings; Life Table Analysis; Logistic Regression Analysis (Unconditional and Conditional); Wilcoxon-Mann-Whitney Test and Related Procedures ; Extended Mantel-Haenszel Procedure: Trend Analysis; Multiple Matched Controls; Correcting for Misclassification in 2 x 2 Tables; Analysis of Paired Samples ; Poisson Probability: Observed vs Expected Events; Poisson Regression Analysis; Power of a Test Comparing Two Proportions or Means; Probability and Inverse Probability Values: Z, t, Chi Square, F; Procedures using Random Numbers; Association Between Ordinal-Scale Variables; Comparison of Two Rates or Proportions; Comparison of Person-Time Incidence Rates; ower and Sample Size for Regression and Correlation Analyses; Comparison of Several Related Samples; Sample Size for Estimation of Proportion, Rate, or Mean; Sample Sizes for Comparison of Two Samples ; Internal Consistency of a Scale; Screening and Diagnostic Tests ; Seasonal Variation ; Smoothing of Curves and Median Polish Procedure; Kaplan-Meier Life Table Analysis, Log-rank and Logit-rank Tests; Calculation of Elapsed Time; Trend Analysis and Multiple Comparisons, and two special calculators: WHATIS and WHATS.

TETRAD (from the TETRAD Project at CMU) -- a free program for creating, simulating data from, estimating, testing, predicting with, and searching for causal/statistical models of categorical (or ordinal) data and to linear models ("structural equation models') with a Normal probability distribution, and to a very limited class of time series models. Provides sophisticated methods in a friendly interface. It performs many of the functions in commercial programs such as Netica, Hugin, LISREL, EQS and other programs, and many discovery functions these commercial programs do not perform. TETRAD is limited to models The TETRAD programs describe causal models in three distinct parts or stages: a picture, representing a directed graph specifying hypothetical causal relations among the variables; a specification of the family of probability distributions and kinds of parameters associated with the graphical model; and a specification of

increments, sequential sets), Logistic regression (1 or 2 continuous variables, 5-level categorical), Survival analysis (accrual options , constant or variable hazard, attrition), Equivalence tests (proportions or means).

StudySize (free demo subset and 30-day trial versions) -- for power / precision / sample size calculations, hypothesis testing, point estimation and confidence interval calculations. Includes equivalence- and non-inferiority testing for most tests, Monte Carlo simulation for small samples; group sequential interim analyses . Tests: Wilcoxon, Mann-Whitney, Sign; Student t, 1-way ANOVA, Fisher's exact, McNemar's,Chi2, Life table, Log-rank,Bioequivalence, Correlation and Regression coefficients. Point estimates & confidence intervals for means, medians, SDs, location & scale params in normal, log-normal,

the numerical values of those parameters.

EasySample -- a tool for statistical sampling. Supports several types of attribute and variable sampling and includes a random number generator and standard deviation calculator. Has a consistent, easy-to-use interface. Results may be saved or read in CSV (spreadsheet compatible) or XML (Internet compatible) file formats or printed.

EpiData -- a comprehensive yet simple tool for **documented data entry**. Overall frequency tables (codebook) and listing of data included, but no statistical analysis tools.

Calculate sample size required for a given confidence interval, or confidence interval for a given sample size. Can handle finite populations. Online calculator also available.

Grocer -- a free econometrics toolbox that runs under Scilab. It contains: most standard econometric capabilities: ordinary least squares, autocorelated models, instrumental variables, non linear least squares, limited dependent variables, robust methods, specification tests (multicolinearity, autocorelation, heteroskedasticity, normality, predictive failure,...), simultaneous equations methods (SUR, two and three stage least squares,...), VAR, VECM, VARMA and GARCH estimation, the Kalman filter and time varying parameters estimation, unit root tests (ADF, KPSS,...) and cointegration methods (CADF, Johansen,...), HP, Baxter-King and Christiano-Fitzgerald filters. It also contains some rare -and useful- features: a *pc-gets* device that performs automatic general to specific estimations, and a *contributions* device, that provides contributions of exogenous variables to an endogenous one for any dynamic equation. Has a -rough- interface with Excel and unlike Gauss or Matlab, it deals with true timeseries objects.

exponential,
binomial, hypergeometric,
Poisson distributions, and
more.

Design-Ease and Design-Expert -- two programs from Stat-Ease that specialize in the design of experiments. Full-function 45-day evaluation copies of both programs are available for download.

AGREE -- to measure agreement of nominal data, where two or more judges classify objects into nominal scale categories.

Bayesware Discoverer -- a computer program able to learn Bayesian Belief Networks from (possibly incomplete) databases. Based on a new estimation method called Bound and Collapse. This is a commercial product, available free for educational and other non-commercial use. See also the freeware product: BKD:

Biomapper -- a kit of **GIS and statistical tools** designed to build **habitat suitability (HS) models and maps** for any kind of animal or plant. Deals with: preparing ecogeographical maps for use as input for ENFA (e.g. computing frequency of occurrence map, standardisation, masking, etc.); Exploring and comparing them by mean of descriptive statistics (distribution analysis, etc.); Computing the Ecological Niche Factor Analysis and exploring its output; and Computing and evaluating a Habitat Suitability map

ROC Curves -- a set of downloadable programs and Excel spreadsheets to calculate and graph various kinds of ROC (Receiver Operator Characteristic) curves.

BKD: Bayesian Knowledge Discoverer -- a computer program able to learn Bayesian Belief Networks from (possibly incomplete) databases. Based on a new estimation method called Bound and Collapse. Developed within the Bayesian Knowledge Discovery project. See also the commercial product, called Bayesware Discoverer, available free for non-commercial use.

RoC: The Robust Bayesian Classifier -- a computer program able to perform supervised Bayesian classification from incomplete databases, with no assumption about the pattern of missing data. Based on a new estimation method called Robust Bayesian Estimator. Developed within the Bayesian Knowledge Discovery project.

DQO-PRO -- a sample-size calculator for MS Windows that performs three types of calculations:

- determining the rate at which an event occurs (confidence levels versus numbers of false positive or negative conclusions),
- determining an estimate of an averge within a tolerable error (given the

Bayesian Knowledge Discoverer.

ZeroRejects -- Implements the "Six Sigma" statistical process control methodology developed by Motorola. The alpha and beta version are freely downloadable. (Win 95/98/NT).

Prognosis -- for analysis of time-series data. Uuses artificial intelligence and powerful statistical methodology to achieve high forecasting accuracy. Easy to use; does not require any background in statistics or time series analysis. Free evaluation copy available for download.

standard deviation of individual measurements), and
- determining the sampling grid necessary to detect “hot spots” of various assumed shapes.

Binomial Probability Program (BPP) is a menu driven program which performs a variety of functions related to the success/ failure situation. Given the probability of occurrence for a specific event, this program calculates the probability that EXACTLY, NO MORE THAN, or AT LEAST a certain number of events occur in a given number of trials for all possible outcomes, and will generate plots for each of these.

The program allows the user to repeatedly combine probabilities in series or in parallel, and at any time will show a trail of the calculations which led to the current probability value. Other program capabilities are the calculation of probabilities from input data, Gaussian approximation, and the generation of a mean time between failure (MTBF) table for various levels of confidence. Up to 2200 trials may be run, limited by IBM PC BASIC memory utilization. It is assumed that the user is familiar with the theory behind binomial probability distribution.

ADE-4 -- multivariate analysis and graphical display software package for Mac andWin 95/NT. Includes component analysis and correspondence analysis, spatial data analysis methods (analogous to Moran and Geary indices), discriminant analysis and within/between groups analyses, many linear regression methods including lowess and polynomial regression, multiple and PLS (partial least squares) regression and orthogonal (principal component) regression, projection methods like principal component analysis on instrumental variables, canonical correspondence analysis and many other variants, coinertia analysis and the RLQ method, and several three-way table (k-table) analysis methods. Graphical displays include an automatic collection of elementary graphics corresponding to groups of rows or to columns in the data table,

automatic k-table graphics and geographical mapping options, searching, zooming, selection of points, and display of data values on factor maps. Simple and homogeneous user interface.

Weibull Trend Toolkit -- Fits a Weibull distribution function (like a normal distribution, but more flexible) to a set of data points by matching the skewness of the data. (Windows)

TURNER -- Macintosh software for interactivly analysing multidimensional discrete data. Uses interactive paradigms from exploratory graphical data analysis to the concise treatment of categorical data, typically arranged in two- or multi-way contingency tables. Including standard features for categorical data like Pearson's chi-squared test and log-linear models it offers the whole goodness-of-fit family of power divergence statistics and the N-value. Interactive contingency tables provide the user with the facility of easily switching between all two-dimensional views of multivariate data. All displays dealing with the same data set are fully linked and may be interacted with directly.

BUGS -- Bayesian inference Using Gibbs Sampling. Software for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods. Command-line interface versions available for major computer platform; a Windows version, WinBUGS, supports a graphical user interface, on-line monitoring and convergence diagnostics.

MSBNx -- a component-based Windows application for creating, assessing, and evaluating Bayesian Networks, created at Microsoft Research. Includes complete help files and sample networks. Bayesian Networks are encoded in an XML file format.

QUEST (Quick, Unbiased and Efficient Statistical Tree), and CRUISE

(Classification Rule with Unbiased Interaction Selection and Estimation. Two statistical decision tree algorithms for classification and data mining, by Wei-Yin Loh and Yu-Shan Shih.

AMELIA -- A program for substituting reasonable values for missing data (called "imputation")

A collection of MS-DOS program from the Downloads section of the QuantitativeSkills web site:

- Hypergeometric -- calculates the hypergeometric probability distribution to evaluate hypothesis in relation to sampling without replacing in small populations
- Binomial -- calculates probabilities for sampling with replacing in small populations or without replacing in very large populations. Can be used to approximate the hypergeometric distribution. The binomial is probably the best known discrete distribution.
- Poisson -- calculates probabilities for samples which are very large in an even larger population. Is used to approximate the binomial distribution, try to compare it with the binomial! The distribution is more often used in a completely different way, for the analysis of how rare events, such as accidents, cumulate for a single individual. For example, you can use it to estimate your chances of getting one, two, three or more accidents in any one year considering that on average people get 'U' accidents per year.
- Negative binomial -- Also used to study accidents, is a more general case than the Poison, it considers that the probability of getting accidents if accidents clusters differently in subgroups of the population. However, the theoretical properties of this distribution and the possible relationship to real events are not well known.
- Negative binomial -- Another version of the negative binomial, this one is

used to do the marginal distribution of binomials (try it!). Often used to predict the termination of real time events. An example is the probability of terminating listening to a non-answering phone after n-rings.

- Multinomial -- Same as the multinomial above, this one for DOS computers.
- Fisher -- Is used to calculate the exact p-value in 2*2 tables. It is o.k. for one sided testing but not so exact for two sided testing, where there are different theories about how to do it. The sum of small p-values is the most used method, but there does not seem to be a good rationale for that. Use the fisher exact instead of the Chi-square when you have a small value in one cell or a very uneven marginal distribution.
- SPRT -- This method of analysis is not often used, which is a pity because it is actually quite good. It is based on the case of phenomena being observed, tested, or data collected, sequentially in time. The testing or data collection is stopped as soon as some upper or lower limit is crossed of the proportion positive or negative events or outcomes relative to the total number observed. Was originally developed to keep the costs of 'destructive' testing low. Is sometimes used in medical trials to monitor the amount of negative side effects and to decide if the trial should be stopped because the number of side effect is considered unacceptably high.
- Chi-square -- Calculates the Chi-square and some other measures for two dimensional tables
- CASRO -- Calculates response rates according to different procedures. The CASRO (Council of American Survey Research Organizations) procedure is the 'accepted' procedure for surveys.

Curve-fitting & Modeling:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of

	commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
EasyReg (Easy Regression Analysis), by Herman J. Bierens. Incredibly powerful and multi-featured program for data manipulation and analysis. Designed for econometrics, but useful in many other disciplines as well. For Win 98/98/NT4. Compumine Rule Discovery System -- easy to use data mining software for developing high-quality rule based prediction models, such as classification and regression trees, rule sets and ensemble models. This program is licensed under the P3 license model wich means that it is free to use forever for developing rule-based predictive models, and can be freely downloaded here. gretl -- a cross-platform (Linux, Windows, Mac, etc.) package for econometric analysis. Has an intuitive interface (English, French, Italian & Spanish). Supports a wide variety of least-squares based estimators, including two-stage & nonlinear least squares, augmented Dickey-Fuller test, Chow test for structural stability, Vector Autoregressions, ARMA estimation. Creates output modelss as LaTeX files, in tabular or equation format. Has an integrated scripting language: enter commands either via the gui or via script, command loop structure for Monte Carlo simulations and iterative estimation procedures, GUI controller for fine-tuning Gnuplot graphs, Link to GNU R for further data analysis. Reads own format XML data files, Comma Separated Values files, Excel and Gnumeric worksheets, BOX1 files, own format binary databases (allowing mixed data frequencies and series lengths) and RATS 4 databases. Includes a sample US macro database. See also the gretl data page. mle - Maximum Likelihood Estimation -- a simple programming language for building and estimating parameters of likelihood models. Originally designed for	CurveExpert -- comprehensive curve fitting system for Windows. Handles linear regression models, nonlinear regression models, interpolation, or splines. Over 30 models built-in; custom user-defined regression models. Full-featured graphing capability. Supports an automated process that compares your data to each model to choose the best curve. 30-day evaluation of shareware package. DTREG generates classification and regression decision trees. It uses V-fold cross-valication with pruning to generate the optimal size tree, and it uses surrogate splitters to handle missing data. A free demonstration copy is available for download.

survival models, but the language has evolved into a general-purpose tool for building and estimating general likelihood models. Available for Windows and Linux; also provides User Manual, Reference Manual, and Quick Reference Card.

WinSAAM -- Windows implementation of SAAM (System Analysis and Modeling Software). Lets you create mathematical models, design and simulate experiments, and analyze data. Models can contain differential equations, which will be numerically integrated and fit to data. Graphic and tabular output is provided.

Boomer -- Non-linear Regression Program for Analysis of Pharmacokinetic and Pharmacodynamic Data. Includes normal fitting, Bayesian estimation, or simulation-only, with integrated or differential equation models. Allows selection of weighting schemes and methods for numerical integration. Free downloads for Macintosh and Windows; online manual, tutorial, sample data sets.

DEMETRA -- user-friendly interface to TRAMO/SEATS and X-12-ARIMA .

JoinPoint Regression Program (from the National Cancer Institute) -- for the analysis of trends using joinpoint models (where several different lines are connected together at the "joinpoints."). Takes trend data (e.g cancer rates) and fits the simplest joinpoint model that the data allow, using a Monte Carlo Permutation method. Models may incorporate estimated variation for each point (e.g. when the responses are age adjusted rates) or use a Poisson model of variation. In addition, the models may also be linear on the log of the response (e.g. for calculating annual percentage rate change). The software also allows viewing one graph for each joinpoint model, from the model with the minimum number of joinpoints to the model with maximum number of joinpoints.

NLREG performs general nonlinear regression. NLREG will fit a general function, whose form you specify, to a set of data values. A free demonstration copy is available for download.

Partitionator -- a fast recursive partitioning engine that uses a learning set to generate rules by which a dependent variable can be predicted, by optimally splitting continuous predictors. Free 30-day evaluation.

NeuroSolutions -- applies neural network technology to many situations, including regression. Free evaluation version does everything except print or save networks.

LOCFIT -- a software system for fitting curves and surfaces to data, using the local regression and likelihood methods. (from Bell Labs) Runs on various platforms under R or S statistical systems; also

	available as a stand-along package for Win95/98/NT. Origin -- technical graphics and data analysis software for Windows. Includes 3D and contour plotting, FFT filtering; works closely with Excel. 30 evaluation. CART -- Salford Systems flagship decision-tree software, combines an easy-to-use GUI with advanced features for data mining, data pre-processing and predictive modeling.

Biostatistics and Epidemiology:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
OpenEpi Version 2.2.1 -- OpenEpi is a free, web-based, open source, operating-system-independent series of programs for use in public health and medicine, providing a number of epidemiologic and statistical tools. It is written in JavaScript and HTML and	HICAST -- a PC-based program for rapid entry of clinical and laboratory

operates similar to a calculator. OpenEpi can be thought of as an important companion to Epi Info, EpiData, SAS, SPSS, and Stata.

M.D. Anderson Statistical Software Library-- A large collection of free statistical software (almost 70 programs!) from the Biostatistics and Applied Mathematics department of the M.D. Anderson Cancer Center. Software is distributed in the form of program source files and/or self-extracting archives of executable programs for Windows, Mac, Unix/Linux environments.

Lifetables -- Windows program for Mortality Analysis for Demography and Epidemiology. The program will calculate the life expectancy, including all intermediary statistics, variance an confidence interval for the life expectancy, Potential Gains in Life Expectancy (PGLE), Years of Potential Life Lost (YPLL) and Lifetime Years of Potential Life Lost (LYPLL). YPLL can be calculated adjusted for competing causes of mortality and both YPLL and LYPL can also be discounted. Two populations can be compared using direct and indirect standardization, the SMR and CMF and by comparing two lifetables. Confidence intervals and statistical test are provided. There is an extensive helpfile in which everything is explained. From the Downloads section of the QuantitativeSkills web site.

Sample Size for Microarray Experiments -- compute how many samples needed for a microarray experiment to find genes that are differentially expressed between two kinds of samples (e.g.: cancer vs. normal tissue), by performing separate gene-by-gene t-tests. You specify how many genes you're looking at, how many false positives you are willing to accept, how large a difference you want to be able to detect (as the fold difference between the two kinds of samples), the power of the test (% of differentially expressed genes likely to be detected by the experiment), and an estimate of the logarithmic SD of the gene intensities.

MIX (Meta-analysis with Interactive eXplanations) -- a statistical add-in for Excel 2000

parameters needed for the calculation of ten internationally applied scoring systems used on the an Intensive Care Unit. Allows sharing of relvant data, so multiple enrties of the same data are not necessary.

or later (Windows only). Ideal for learning meta-analysis (reproduces the data, calculations, and graphs of virtually all data sets from the most authoritative meta-analysis books, and lets you analyze your own data "by the book"). Handles datasets with dichotomous & continuous outcomes; calculates Risk Diff, RR, OR, Mean Diff, Hedges's g, Cohen's d; performs standard & cumulative meta-analysis with CI ,z & p; fixed and random effects modeling; Cochran's Q with p-value; Higgins's I2 and H with CI; and publication bias tests: Rank correlation (tau-b) test with z & p, Egger's and Macaskill's regression tests with CI, and Trim-and-Fill. Generates numerous plots: tandard and cumulative forest, p-value function, four funnel types, several funnel regression types, exclusion sensitivity, Galbraith, L'Abbe, Baujat, modeling sensitivity, and Trim-and-Fill.

EWOC - Escalation With Overdose Control -- a Bayesian method for selecting dose levels in Phase I Clinical Trials while controlling the probability of exceeding the maximum tolerated dose. This is a stand-alone Windows (95 through XP) program that receives information about dose-limiting toxicities (DLTs) observed at some starting dose, and calculates the doses to be administered next. DLT information obtained at each dosing level guides the calculation of the next dose level. (For some strange reason, the EWOC download web site does not work properly with the FireFox web browser; but it does work with MS Internet Explorer.)

STPLAN -- Performs power, sample size, and related calculations needed to plan studies. Covers a wide variety of situations, including studies whose outcomes involve the Binomial, Poisson, Normal, and log-normal distributions, or are survival times or correlation coefficients. Available for MS-DOS and Mac; also as Fortran and C source code.

Epi InfoVersion 3.5.1 -- Public domain statistical software for epidemiology developed by Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia (USA). Epi Info has been in existence for over 20 years and is currently available for Microsoft

Windows. The program allows for data entry and analysis. Within the analysis module, analytic routines include t-tests, ANOVA, nonparametric statistics, cross tabulations and stratification with estimates of odds ratios, risk ratios, and risk differences, logistic regression (conditional and unconditional), survival analysis (Kaplan Meier and Cox proportional hazard), and analysis of complex survey data. The software is in the public domain, free, and can be downloaded from http://www.cdc.gov/epiinfo. Limited support is available.

PEPI -- a collection of 43 small DOS / Windows programs that perform a large assortment of statistical tests. They can be downloaded individually, or as a single ZIP file.

Free Public Health & Epidemiology Software -- written by Mark Myatt)and others:

- WINPEPI (PEPI for Windows) -- Windows versions of the renowned PEPI suite of programs. A freely-accessible article describing the new features of WinPEPI can be accessed here.
- EpiCalc 2000 -- Statistical calculator
- FP Advisor -- Foodborne disease database
- SigmaD -- Standardisation of measurement
- SOUNDEX Calculators -- Confidentiality of data / identification of duplicate records
- SampleXS -- Sample size calculator for cross-sectional surveys
- SampleLQ -- Sample size calculator for LQAS surveys
- SampleRate -- Sample size calculator for a single rate
- EpiGram -- Simple diagramming software
- Statistical Utilities -- Miscellaneous statistical and epidemiological utilities (by Ray Simons and others)
- EpiInfo PLUS -- A version of the classic DOS version of EpiInfo
- EpiInfo Add-ins -- Logistic regression and survival analysis for EpiInfo .REC files

- MUAC screening tool -- Software to determine MUAC cut-points for two-stage screening in nutritional emergencies.
- CSAS coverage calculator -- An Excel spreadsheet for calculating coverage estimates and drawing plots and maps from CSAS survey data. Also provides capture-recapture estimates of the sensitivity of a case-finding procedure. A spreadsheet containing example data is available here.
- A brief introduction to using the R environment for analysing epidemiological data in Adobe Acrobat (.PDF) format with sample data, in a ZIP file here.

PAMCOMP (Person-years And Mortality COMputation Program) -- a free Windows 95/98/NT application for calculating person-years and standardised mortality ratios (SMRs). The calculation of person-years allows flexible stratification by sex, and self-defined and unrestricted calendar periods and age groups, and can lag person-years to account for latency periods.The SMR computation includes calculation of 90%, 95%, and 99% confidence intervals. Has filters for ASCII, dBase, Excel, Access, Paradox to import cohort and reference data and to export distributions of person-years and deaths.

ARIMA -- a seasonal adjustment program for PC and Unix, developed by the Census Bureau.

DEMETRA -- (Win 9x/NT) a user-friendly interface to the seasonal adjustment methods TRAMO/SEATS and X-12-ARIMA . Developed by Eurostat to facilitate the application of these modern time series techniques to large-scale sets of time series and in the explicit consideration of the needs of production units in statistical institutes. Client/server architecture can access various kinds of databases and files. Contains two main modules: seasonal adjustment and trend estimation with an automated procedure (e.g. for unexperienced users or for large-scale sets of time series), and with a user-friendly procedure for detailed analysis of single time series.

Meta-analysis 5.3 -- Free DOS statistics software for meta-analysis. Probably still the most frequently used meta-analysis software in the world. Can select the analysis of exact p values or effect sizes (d or r, with a cluster size option). Can plot a stem-and-leaf display of correlation coefficients. A utility menu is provided that allows various transformations and preliminary computations that are typically required before the final meta-analysis can be performed.

EasyMA -- a free user-friendly MS-DOS program for the meta-analysis of clinical trials results. Developed to help physicians and medical researchers to synthesize evidence in clinical or therapeutic research.

EPIMETA (from CDC) -- a DOS-based meta-analysis program that features a Windows-like interface which makes data entry, file manipulation, and subgroup analysis easy.

Life Table -- available in Lotus and Excel formats.

ABSRISK -- a program (MS-DOS) for estimating absolute risks from relative risks. Uses age-specific mortality and morbidity data to convert relative risk estimates into absolute risk estimates. That is, it estimates the probability that a patient will suffer a specific morbid or mortal outcome in a given time interval. The user first specifies a data file that contains the needed mortality and morbidity data for the disease of interest. She then gives her patient's age and relative risk, and the time interval over which the risk estimate is to be derived. The program derives this risk, which is given both interactively and in a log file.

Biodiversity Research Software -- Five software packages, with documentation:

1. LUMP, LINK, and JOIN: Utility Programs for Biodiversity Research
2. COLLECT1 and COLLECT2: Programs for Calculating Statistics of Collectors'

Curves
3. BOUNDARY: A Program for Detecting Boundaries in Ecological Landscapes
4. EXTSPP1 and EXTSPP2: Programs for Comparing and Performance-Testing Eight Extrapolation-Based Estimators of Total Taxonomic Richness
5. RARE, SPPDISS, and SPPRANK: Programs for Detecting Between- Sample Differences in Community Structure

Surveys, Testing, and Measurement:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
CCOUNT -- a package for market research data cleaning, manipulation, cross tabulation and data analysis. Similar to, and uses the same syntax as, SPSS-MR "Quantum", a well known commercial package for processing market research data. Available for Windows, Linux, and SunOS. C++ source code also available, under the GNU General Public Licerse. ProtoGenie -- a free extensible web-based environment for research design and data collection for surveys, experiments, clinical trials, time series, cognitive and vision research, and methods courses. Lets you specify groups and define measurement and treatment events and their sequencing. The goal is to let users move smoothly from research design and data collection to interim and final statistical analysis.	Form Artist -- lets you design and create online forms for data collection via the Web. Forms and surveys run on any web server (Microsoft, Unix, Linux), and work with all browsers (no plugins required. WYSIWYG interface gives complete control over the appearance of forms (any shape, size, number of pages, color scheme). Create multi-page forms on the same web page without reloading.. Supports the usual data entry fields (text, numbers, lists, checkboxes etc.), also unique objects such as picture grids and emoticons. Can fill in forms online or offline. Completed data can then be sent back via email or by file. Free evaluation version available. AssiStat -- a Windows-based package of calculations and analyses useful in educational and psychological research, practice, and in measurement and statistics

GGUM2004 (Item Response Theory Models for Unfolding) -- a Windows-based program that estimates parameters in the generalized graded unfolding model (GGUM; Roberts, Donoghue, & Laughlin, 2000). Has a user-friendly interface to prepare command files, run the core estimation program, and display results. Allows different questionnaire items to have varying numbers of response categories (useful when sparse responses require recoding into fewer response categories. Handles sporadically missing responses. Provides item fit statistics and diagnostic graphics of performance.

Rasch Measurement Software -- deals with the various nuances of constructing optimal rating scales from a number of (usually) dichotomous measurements, such as responses to questions in a survey or test. Several free student/demo software packages are available. These may be freely downloaded, used, and distributed, and they do not expire. They are:

- BIGSTEPS -- a DOS-based precursor to the Windows-based WINSTEPS Rasch measurements program.
- MINISTEP -- a free evaluation/student version of WINSTEPS. It has complete WINSTEPS functionality, but is limited to 25 items and 100 persons (cases).
- MINIFAC -- a free evaluation/student version of FACETS (Many-Facet Rasch Analysis). Contains all features except limited to 2,000 data points

courses. Designed as a complement to typical statistical packages rather than as a primary analysis tool, it picks up where primary analysis packages usually fall short--in performing secondary analyses like correction of correlations for restriction in range or less-than-perfect reliability, and other specialized analyses and calculations usually not available in standard packages without special programming. Free demo available.

StatPac Survey Software -- to design andimplement surveys, and to acquire, manage and analyze data from surveys. Supports multiply data types and question formats, multi-language spell-checking, large files (2,000 variables & 10,000,000 cases, basic statistics (crosstab & banner tables) & graphics, automatic coding of text responses, and data import / export capabilities. Optional Web Survey Module and Advanced Statistics Module (curve fitting, multiple regression, logistic regression, factor, analysis of variance, discriminant function, cluster, and canonical correlation). A demo version is available (limited to 35 cases).

NewMDSX -- software for Multidimensional Scaling (MDS), a term that refers to a family of models where the structure in a set of data is represented graphically by the relationships between a set of points in a space. MDS can be used on a variety of data, using different models and allowing different assumptions

(responses).

Q-Method -- a statistical program for analyzing data from the Q-Sort Technique. Enter data (Q-Sorts) the way they are collected, i.e. as 'piles' of statement numbers. It computes intercorrelations among Q-Sorts, which are then factor-analysed with the Centroid (or, alternatively, PCA) method. Resulting factors can be rotated either analytically (Varimax), or judgmentally with the help of two-dimensional plots. Finally, after selecting the relevant factors and 'flagging' the entries that define the factors, the analysis step produces an extensive report with a variety of tables on factor loadings, statement factor scores, discriminating statements for each of the factors as well as consensus statements across factors, etc.

AnSWR -- Analysis Software for Word-based Records -- a free software system from the C.D.C. for coordinating and conducting large-scale, team-based analysis projects that integrate qualitative and quantitative techniques (for Windows).

ez-text -- a software program from the C.D.C. developed to assist researchers create, manage, and analyze semi-structured qualitative databases.

CSPro (Census and Survey Processing System) -- a public-domain software package for entering, tabulating and mapping census and survey data.

about the level of measurement. This site offers a free month trial of the Windows version; a completely free copy of constituent programs, notes, documentation, test Input& Output in MS-DOS; a not-for-profit full Windows copy priced at cost; and a site with a range of data, cross-reference, & information.

GLIMMIX -- a powerful approach to segmentation based on latent class models. Analysis of brand choice, purchase frequency and preference data.

CORWIN -- a program for correspondence analysis, which decomposes relations in a two-way table.

ConTEST -- a decision support system for assembly of educational and psychological tests from item banks.

MUDFOLD (Multiple UniDimensional unFOLDing) -- for analyzing proximity data (e.g., attitudes, preferences, or choices) with the Coombsian unfolding model.

WINMIRA -- Latent Class Analysis (LCA), the Rasch model (RM), and the Mixed Rasch model (MRM) and Hybrid models (HYBRID).

T-Rasch -- exact or non-parametric tests for the Rasch model.

IMPS (Integrated Microcomputer Processing System) -- performs the major tasks in survey and census data processing: data entry, data editing, tabulation, data dissemination, statistical analysis and data capture control. (from CDC)

WebQ -- a set of HTML files for performing Q-Sorts online and collecting the data for subsequent analysis.

Stats -- Windows program for several commonly-needed statistical functions for marketing researchers: random numbers;sample sizes needed for surveys; mean, standard deviation, standard error and range for keyboard-entered data; standard error of a proportion; significance testing between two percentages from independent samples; significance between two percentages from dependent samples; significance testing between two averages from independent samples; contingency table analysis (i.e., Chi-Square)

SABRE -- for the statistical analysis of multi-process random effect response data. Responses can be binary, ordinal, count and linear recurrent events; response sequences can be of different types. Such multi-process data is common in many research areas, e.g. the analysis of work and life histories. Sabre has been used intensively on many longitudinal datasets surveys either with recurrent information collected over time or with a clustered sampling scheme.

POSDEM -- Uses simulation techniques to analyze and

LPCM-WIN -- a menu-driven program to apply 'Linear Partial Credit Models' in item analysis and measurement of change.

Kwalitan -- for analysis of qualitative data, such as protocols of interviews, articles, and annual reports.

compare alternate sampling strategies for surveys. Performs power / sample size / precision analyses for different sampling methods: systematic, stratified, random, etc. Windows versions available in Spanish and English.

WISC-III Profile Calculator for Macintosh and Windows -- uses generalized distance method to determine if the subtest profile of a single case is multivariately unusual or common in comparison to subtest clusters found in the WISC-III standardization sample. (Mac, 360K; Win anticipated in September)

DEMETRA -- user-friendly interface to TRAMO/SEATS and X-12-ARIMA .

Sociological Insights -- displays statistical information in an easy-to-use format, designed for teaching quantitative sociological reasoning. It uses aggregate data from the 50 U.S. states to teach the principles of distribution, correlation, and regression. It uses questionnaire data from the 2000 and 1994 General Social Surveys to teach distribution and cross-tabulation. The States module has 289 variables in all. The Survey module displays 249 variables from the 2000 GSS, plus (as a separate data set) 113 variables from the 1994 GSS.

Excel Spreadsheets and Add-Ins:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial

	packages; can be freely downloaded, but are usually restricted or limited in some way.
PopTools -- Windows DLL for Excel 97 and 2000 (PC's only). Facilitates analysis of matrix population models & simulation of stochastic processes. Adds a new menu item and installs many powerful functions: matrix decompositions (Cholesky, QR, singular values, LU), eigenanalysis (eigenvalues and eigenvectors of square matrices) and formulas for generation of random variables (Normal, binomial, gamma, exponential, Poisson, logNormal). Also has routines for iterating spreadsheets to run Monte Carlo simulations, conduct randomisation tests (including the Mantel test) and calculate bootstrap statistics. Some facilities for maximum-likelihood parameter estimation, and some other generally useful functions. Free download from website, which also has documentation, examples, and related links. SimulAr -- Provides a very elegant point-and-click graphical interface that makes it easy to generate random variables (correlated or uncorrelated) from twenty different distributions, run Monte-Carlo simulations, and generate extensive tabulations and elegant graphical displays of the results. EZAnalyze -- enhances Excel (Mac and PC) by adding "point and click" functionality for analyzing data and creating graphs (no formula entry required). Does all basic "descriptive statistics" (mean, median, standard deviation, and range), and "disaggregates" data (breaks it down by categories), with results shown as tables or disaggregation graphs". Advanced features: correlation; one-sample, independent samples, and paired samples t-tests; chi square; and single factor	XLStatistics -- a set of Excel (ver 5+) workbooks for statistical analysis of data. A step-by-step guide to data analysis with separate workbooks for handling data with different numbers and types of variables. Contains most standard analyses, analyses using only summary data, power / sample size , nonparametrics, curve fitting , non-linear regression, analysis for 2x2 tables. XLStatistics is not an Excel add-in and all the working and code is visible. A free version for analysis of 1- and 2-variable data is available. XLMathematics -- A set of Excel (Ver 5+) for mathematical computations: graphing , calculus (computing limits, computing and graphing derivatives and/or tangent lines, evaluating integrals using various techniques), Linear algebra (Gauss-Jordan elimination, allowing step-by-step views). Analyse-it -- includes over 30

ANOVA.
Update Available! The latest version can create z-scores, percentile ranks, and random numbers as new variables; has repeated-measures ANOVA; does simple post hoc tests for single factor and repeated-measures ANOVA; can graph multiple variables on a single graph, and can add error bars for +/- 2 SD's; adds the *sum* function to the *disaggregate* and *descriptive statistics* functions, and the *mode* function to *descriptive stats*; adds *delete sheets*; adds English & Spanish language options, and works better in international environments; incorporates various bug fixes; and contains an updated user manual.

EZ-R Stats -- supports a variety of analytical techniques, such as: Benford's law, univariate stats, cross-tabs, histograms. Also supports databases such as mySQL, SQLite, MS-Access, MS-SQL. Simplifies the analysis of large volumes of data, enhances audit planning by better characterizing data, identifies potential audit exceptions and facilitates reporting and analysis. This language is a Computer Assisted Audit Technique (CAAT) in support of COSO, SAS 78, SAS 99 and analysis required by Sarbanes-Oxley.

SSC-Stat -- an Excel add-in designed to strengthen those areas where the spreadsheet package is already strong, principally in the areas of data management, graphics and descriptive statistics. SSC-Stat is especially useful for datasets in which there are columns indicating different groups. Menu features within SSC-Stat can:

- help users manipulate their data (stacking, unstacking columns, 2-way unstacking, lookups, generating factors, etc.);
- generate good graphs (X-Y Scatter Plot, Category-Value Plot, Boxplot, Normal Probability Plot, Density Estimate), that can be

parametric & non-parametric statistical functions, including multiple linear regression analysis, ANOVA, & chi-square statistics. A separate specialized package for clinical method evaluation provides NCCLS and IFCC procedures for accuracy & imprecision.

Statistical Process Control (SPC) and Reliability spreadsheets from John Zorich's web site -- designed to simplify activities in Production and R&D. Formally validated to be "GMP" and "Part 11" compliant . Demo's of spreadsheets include:

- **Variables Data SPC** -- XbarR, XbarS, XmR, histograms, capability indices, preformatted customizable printable report. Automaticly identify out-of-control points.
- **Count Data SPC** -- P and U SPC charts, pareto chart, preformatted customizable printable report. Sutomaticly identify out-of-control points.
- **Reliability Statistics Basics** -- component reliability using K-factors, stress/strength analysis,

edited and polished like any other Excel graph ;
- provide basic statistical analysis (descriptive statistics, summary statistics, 1- and 2-sample t tests, 1- and 2-sample tests of proportion).

22 Distribution Functions -- There is one spreadsheet for each of the following distribution functions: Beta, Binomial, Chi-Square, Discrete Uniform, Gamma, Geometric, Hypergeometric, Multivariate Hypergeometric, Laplace, Logistic, Multinomial, Negative Binomial, Normal, Bivariate Normal, Log-normal, Pareto, Poisson, Rectangular, Snedecor F, Student-t, Triangular, and Weibull. Each spreadsheet gives a graph of the distribution, along with the value of various parameters, for whatever shape and scale parameters you specify. You can also download a ZIP file containing all 22 spreadsheets.

Sample-size calculator for cluster randomized controlled trials, which are used when the outcomes are not completely independent of each other. This independence assumption is violated in cluster randomized trials because subjects within any one cluster are more likely to respond in a similar manner. A measure of this similarity is known as the intra-correlation coefficient (ICC). Because of the lack of independence, sample sizes have to be increased. This web site contains two tools to aid the design of cluster trials – a database of ICCs and a sample size calculator (along with instruction manuals).

DAG_Stat -- calculates an enormous number of quantities from a 2 -by-2 table:

- for **diagnostic tests**: sensitivity, sensitivity of a random test given the observed prevalence and test level., sensitivity quality index,

failure analyses, for "normally distributed" and unknown distributions. Stress / strength formula has been modified to allow input of a "confidence" level, if desired.
- **Reliability Plotting** -- Component reliability using "Reliability plotting" ("probability plotting", "rectification", etc.). Can confirm normality, or can identify normalizing transformation.
- **Power Curves for t-Tests** -- Power vs. Sample Size, Power vs. Hypothesized Difference, Power vs. Alpha, and Power vs. Population SD.
- **Statistical Analysis of Gages** -- for quantifying measurement uncertainty. Methods include Gage R&R (up to 3 persons, 3 gages, 3 replicates, and 10 parts), Gage Correlation (up to 3 gages), Gage Bias, Gage Linearity, Spec/Inaccuracy Ratios, and Guardbanding..
- **C = 0 Sampling Plans** -- two types of OC curves, and AOQL for chosen plan. Calculates the exact

specificity, specificity of a random test, specificity quality index, efficiency (the correct classification rate), efficiency of a random test, quality index, Youden's index, the predictive value of positive test, predictive value of a positive random test, predictive value of negative test, predictive. value of a negative random test, likelihood ratio of a positive and negative tests, the odds ratio, false positive and false negative rates, prevalence observed in the sample and test level (proportion of subjects classified as 'positive.'

- for **interrater agreement**: Cohen's Kappa, observed agreement, chance agreement, agreement about positive and negative cases, Byrt's bias index, Byrt's prevalence asymmetry index, bias adjusted Kappa, prevalence & bias adjusted Kappa. DAG_Stat also calculates Dice's index, Yule's Q (Gamma), Phi, Scott's agreement index, the tetrachoric correlation coefficient, Goodman & Kruskal's tau, Lambda, the Uncertainty Coefficient, Pearson's Chi Square (with and without Yates' correction), the likelihood ratio Chi Square, McNemar's Test, (with and without Yates' correction).

MIX (Meta-analysis with Interactive eXplanations) -- a statistical add-in for Excel 2000 or later (Windows only). Ideal for learning meta-analysis (reproduces the data, calculations, and graphs of virtually all data sets from the most authoritative meta-analysis books, and lets you analyze your own data "by the book"). Handles datasets with dichotomous & continuous outcomes; calculates Risk Diff, RR, OR, Mean Diff, Hedges's g, Cohen's d; performs standard & cumulative meta-analysis with CI ,z & p; fixed and random effects modeling; Cochran's Q with p-value; Higgins's I2 and H with CI; and publication bias tests: Rank correlation (tau-b) test with z & p, Egger's and Macaskill's regression tests with CI, and Trim-and-Fill. Generates numerous plots: tandard and cumulative forest, p-value function, four funnel types, several funnel regression types,

absolute smallest sample size that gives the desired protection level for a given exact size lot (up to 1000).

exclusion sensitivity, Galbraith, L'Abbe, Baujat, modeling sensitivity, and Trim-and-Fill.

OZGRID -- contains over 4000 pages (and growing) of information on Excel and VBA for Excel. Many add-on's are for sale, but there is also an enormous amount of totally free content: downloads, a free 24/7 question and answer support forum for MS Office, a free Excel monthly newsletter full of detailed tips, tricks, hacks and more for Excel and VBA.

Very-high-precision Statistical Probability Functions -- Provides double-precision (16 significant figures) mass , density, cumulative, inverse probability distributions, critical values, and confidence bounds for the geometric, negative binomial, binomial, Poisson, hypergeometric, negative hypergeometric, exponential, normal, chi-square, gamma, Student t, Fisher F and beta; non-central gamma, chi-square, beta, t and F; and the mixed Gamma-Poisson, Beta-Binomial, and Beta-Negative-binomial distributions. The routines are programmed in VBA, embedded within an Excel spreadsheet that illustrates the usage of each of them.

DE Histograms -- an Excel add-in that provides comprehensive descriptives stats, histograms, outlier detection, normality testing, and much more.

Exact confidence intervals for samples from the Binomial and Poisson distributions -- an Excel spreadsheet with several built-in functions for calculating probabilities and confidence intervals. (42k long).

BiPlot -- by Ilya Lipkovich and Eric P. Smith, of Virginia Tech. A user-friendly add-in for Excel to draw a biplot display (a graph of row and column markers from data that forms a two-way table) based on results

from principal components analysis, correspondence analysis, canonical discriminant analysis, metric multidimensional scaling, redundancy analysis, canonical correlation analysis or canonical correspondence analysis. Allows for a variety of transformations of the data prior to the singular value decomposition and scaling of the markers following the decomposition.

Statistical Process Control (SPC) and Reliability spreadsheets from John Zorich's web site -- designed to simplify activities in Production and R&D. Formally validated to be "GMP" and "Part 11" compliant . Free spreadsheets include:

- **Self-made Sampling Plans** -- Examine the OC curves for your own custom sampling plans. Use either binomial or hypergeometric calculations. Now be able to explain the "valid statistical rationale" of the sampling plans you already use.
- **Sequential Sampling Plans** -- Provides an analysis and planning tool for sample sizes in situations where lots undergo sequential inspections (e.g., 1st by Manufacturing, 2nd by QC, and finally by QA).

Lifetable -- does a full abridged current life table analysis to obtain the life expectancy of a population. Furthermore, one can calculate Potential Gains in Life Expectancy (PGLE) after removing cause k, considering competing causes of death; the (Premature) Years of Potential Life Lost (YPLL), this is the number of person years added to the total number of person years lived in a population if cause of death k would be removed; the Standardized Mortality Ratio (SMR), standardized numbers per 100,000 and the Comparative Mortality Figure (CMF) can also be calculated. From the Downloads section of the QuantitativeSkills web

site.

Intracorrelation -- does intra correlation calculations for dichotomous or binary yes/no type outcome variables according to two different methods proposed for the single cluster one by Fleiss and another one by Bennett et.al. A third spreadsheet concerns a method for two clusters by Donner and Klar. You will have to insert your own data by overwriting the tables in the second (total number of positive responses) and third (total number of negative responses) or fourth column (total number). From the Downloads section of the QuantitativeSkills web site.

Weighted Least Squares Linear Fits -- an Excel add-in from Philip Kromer (Univ. of Texas)

Programming Languages and Subroutine Libraries:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
MuPAD -- a very powerful and general computerized algebra system, developed at the University of Paderborn, now distributed by SciFace Software. In the same category as Mathematica and Maple, it does numerical calculations, symbolic manipulation (algebra, differentiation & integration), graphing, and programming. A free "lite" (but still very powerful) version for PC and Mac can be downloaded. Statistics101 -- executes programs written in the easy-to-learn **Resampling**	Resampling Stats -- a different approach to learning statistics and performing statistical analyses, using simulation with random numbers instead of complex mathematics. 30-day trial version available for Win 95/NT.

Stats statistical simulation language. You write a short, simple program in the language, describing the process behind a probability or statistics problem. Statistics101 then executes your Resampling Stats model thousands of times, each time with different random numbers or samples, keeping track of the results. When the program completes, you have your answer. Runs on Windows, Mac, Lunux -- any system that supports Java.

R -- a programming language and environment for statistical computing and graphics. Similar to S or S-plus (will run most S code unchanged). Available for Windows, various Unix flavors (including Linux), NextStep and Mac. Provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. Well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. The R environment includes:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

ILNumerics -- a numerical library for .NET that turns C# into a 1st class mathematical language. It offers both scientists and software developers convenient syntax (similar to Matlab), toolboxes for statistical functions and machine learning, high performance, wide platform support and 2D and 3D

O-Matrix -- an extensive matrix manipulation system (for Windows) with lots of statistical capability. The "Light" version can be freely downloaded. Some capabilities include:

- **Matrix Functions**: determinant, eigenvalues and eigenvectors, systems of equations
- **Statistics**: minimum, maximum, mean, median, standard deviation, linear regression, correlation, covariance, sorting, t-distributions, f-distributions, probability, normal distributions, population simulations, Kolmogorov-Smirnov Test
- **Optimization**: linear & nonlinear least squares, with and without box constraints and with or without derivatives, quadratic and general nonlinear programming, linear complementarity

visualization features. There's a free "Community" edition and a pay-for "Professional" edition. Both have the same features and capabilities; they differ in how you would re-distribute them in your own software products.

Zelig -- an add-on for R that can estimate, help interpret, and present the results of a large range of statistical methods. It translates hard-to-interpret coefficients into quantities of interest; combines multiply imputed data sets to deal with missing data; automates bootstrapping for all models; uses sophisticated nonparametric matching commands which improve parametric procedures; allows one-line commands to run analyses in all designated strata; automates the creation of replication data files so that you (or anyone else) can replicate the results of your analyses (hence satisfying the replication standard); makes it easy to evaluate counterfactuals; and allows conditional population and superpopulation inferences. It includes many specific methods, based on likelihood, frequentist, Bayesian, robust Bayesian, and nonparametric theories of inference. Zelig comes with detailed, self-contained documentation that minimizes startup costs for Zelig and R, automates graphics and summaries for all models, and, with only three simple commands required, generally makes the power of R accessible for all users. Zelig also works well for teaching, and is designed so that scholars can use the same program with students that they use for their research.

Apophenia -- a statistics library for C. It provides functions on the same level as those of the typical stats package (OLS, probit, singular value decomposition, &c.) but doesn't tie the user to an ad hoc language or environment.

Octave -- a high-level mathematical programming language, similar to MATLAB, for numerical computations -- solving common numerical linear algebra problems, finding the roots of nonlinear equations, integrating ordinary functions, manipulating polynomials, and integrating ordinary differential and differential-

problems

- **Random Simulations**: uniform and normal random number generators, auto-regressive process simulation
- **Special Functions**: error, gamma, incomplete beta, Y and J Bessel
- **Also**: quadrature, differential equations, Fourier analysis, spectral estimation, convolution, FFT, Interpolation, filtering, Kalman-Bucy filtering, wWavelets: Haar and Daubechies transforms, polynomials, and general functions (trig, hyp, inv trig & hyp, exp, log, roots, forward & backward difference approximations to the derivatives of vector-valued functions

Also provides extensive plotting capabilities, with multiple windows, axis scaling & labeling,

algebraic equations. It is easily extensible and customizable via user-defined functions written in Octave's own language, or using dynamically loaded modules written in C++, C, Fortran, or other languages. Runs under Linux and Windows.

J -- a modern, high-level, general-purpose, high-performance programming language. Runs on Windows, Unix, Mac, and PocketPC handhelds. J runs both as a GUI and in a console (command line). Much like *APL*, but uses "conventional" symbols, rather than *APL*'s a specialized character set. J is particularly strong in the mathematical, statistical, and logical analysis of arrays of data. J systems have:

- an integrated development environment
- standard libraries, utilities, and packages
- a form designer for your application forms
- an event-driven graphical user interface to your application
- interfaces with other programming languages and applications
- integrated 2d and 3d graphics
- memory mapped files for high performance data applications

Matvec -- an object oriented programming language with extensive statistical capabilities. Can handle problems ranging from matrix and vector manipulation to the analysis of linear and generalized linear mixed models. Runs in Linux and Windows environments; has a command-line (non-GUI) user interface, and a strong "Unix-like" flavor.

mle - Maximum Likelihood Estimation -- a simple programming language for building and estimating parameters of likelihood models. Originally designed for survival models, but the language has evolved into a general-purpose tool for building and estimating general likelihood models. Available for Windows

titling, free-form text, selectable fonts. Plots exportable to word processors, spreadsheets, etc. Plot Types: line, contour, surface, mesh, bar, stair, polar, vector, error bar, smith charts, and histogram; line plots can contain unlimited points per curve and hundreds of curves per plot; two- and three-dimensional plotting is supported which provides additional flexibility with contours and surface plots; multiple colors, markers, and line types.

and Linux; also provides User Manual, Reference Manual, and Quick Reference Card.

Ox -- an object-oriented matrix programming language with a comprehensive mathematical and statistical function library. Matrices can be used directly in expressions, for example to multiply two matrices, or to invert a matrix. The major features of Ox are its speed, extensive library, and well-designed syntax, which leads to programs which are easier to maintain. Versions of Ox are available for many platforms. The "Console" version can be freely downloaded for academic and research use; the "Professional" version must be purchased.

Mx -- a matrix algebra interpreter and numerical optimizer for exploration of matrix algebra. Many built-in fit fuctions for structural equation modeling and other statistical modeling. Has fitting fuctions like those in LISREL, LISCOMP, EQS and CALIS, along with facilities for maximum likelihood estimation of parameters from missing data structures, under normal theory. Users can easily specify complex 'nonstandard' models, define their own fit functions, and perform optimization subject to linear and nonlinear equality or boundary constraints.

JDB -- Relational Database and Elementary Statistics for a Unix environment. Useful for manipulating experimental data (joining files, cleaning data, reformatting for input into other programs). Computes basic statistics (mean, std. dev., confidence intervals, quartiles, n-tiles, percentiles, histograms, correlations, z-scores, t-scores.

B/D -- an interactive programming language for *a priori* and diagnostic analyses of Bayes linear statistical problems (subjective statistical analyses based on expectation and covariance structures, rather than on distributional assumptions). Quickly and easily specify beliefs about quantities of interest,

attach data to some or all of those quantities, and carry out the general process of Bayes linear adjustment. Produces interactive Bayes linear influence diagrams for the adjustments, providing simple graphical summaries of the adjustments and accompanying diagnostics.

MacANOVA -- comprehensive statistical package for the Mac and PC/Windows. MacAnova has macros which are used just like functions. Several macros are built in, and three files of additional macros (general, time series, design of experiments) are distributed with MacAnova. Like S, MacAnova is a programming language with for and while loops, if, else, elseif, break, and a full range of operations including bit manipulation.

Lisp-Stat -- an extensible statistical computing environment for data analysis, statistical instruction and research, and for exploring the use of dynamic graphical methods. Based on an extended subset of Common Lisp, performs element-wise operations on lists and vectors, and adds a variety of basic statistical and linear algebra functions. Graphics system is object-oriented, and can be customized and adapted. Supports linear and nonlinear regression models and generalized linear models. Runs on Mac, X-window (UNIX), and MS Windows.

Scripts and Macros:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or

	limited in some way.
SPSS Syntax Files -- a large collection of SPSS routines for randomized study design , sampling strategies, meta-analysis, sanple size for confidence intervals, correlation tests, psychometry and other areas. The documentation is in Portuguese, but the scripts are usable as-is. You can have AltaVista automatically translate the page into English by going here, but do **not** use the "translated" scripts! The author has recently added two additional sections (in English) -- one for Dyadic Data Analysis, and one for Simple and Complex Random Assignment for Experimental Designs Link-King -- a SAS program to detect duplicate entries in a file, or to link matching records in two files, based on criteria like names (first, middle, last, maiden, nickname), date of birth, gender, and social security number. A graphical interface, a “Link King for Knaves” feature, and a powerful interface for manually reviewing uncertain matches make it easy to use. It features both probabilistic and deterministic record linkage algorithms, phonetic name matching (NYSIIS and Soundex), and many other features for dealing with "mushy matches".	

Miscellaneous:

Completely Free... can be freely downloaded and used in their fully-functional mode (no strings attached)	**Free, but...** "demonstration" or "student versions" of commercial packages; can be freely downloaded, but are usually restricted or limited in some way.
Text-Stat -- Free Windows program that analyzes ASCI/ANSI texts and HTML files (directly from the internet) and produces word frequency lists and "concordances" (sorted key-word-in-context listings). Can traverse an entire web site, acquiring pages for analysis.	Advanced Grapher (formerly called Serpik Graph) -- a very sophisticated function graphing program - can also plot tables and perform regression. A 30-day full-functioned trial version can be downloaded. CoPlot 6.2 -- for publication-quality 2D and 3D scientific graphs (from data and equations), maps, and technical drawings. From

DeltaStat -- performs statistical calculations on data from 2D gel experiments quantified in Delta2D. Makes use of R and MySQL to perform much faster than the functions provided in Delta2D. Currently provides two sample t-test, a highly configurable database query, multiple analyses per query to analyze proteins that have both higher and lower expression in control versus experimental groups, and support for experiments with variable numbers of control and experimental replicates.

Sampling SIM: Downloadable program (for Mac or Windows) to explore sampling distributions of sample means and proportions. It provides separate windows for building population distributions, drawing and viewing random samples from the population, exploring the behavior of sampling distributions of sample means, and exploring the behavior of confidence intervals.

First Bayes -- a free, easy-to-use Windows application for elementary Bayesian Statistics. Performs most standard, elementary Bayesian analyses, including: plotting and summarizing distributions, defining and examining arbitrary mixtures of distributions, analysis of two kinds of linear model (one or more normal samples

CoHort Software. Creates precise technical drawings using drawing objects, genetic maps, field maps, flow charts, apparatus diagrams, circuit diagrams, chemical structures, etc. Text in drawing objects and graphs can include HTML-like text formatting tags and over 1000 special characters. Supports animated graphs. Exports graphs to .eps, .gif, .jpg, .pdf, .png, .svg, .wmf, and others. Has an auto-recorder and macro programming language. Invoke CoPlot from the command line, batch files, shell scripts, pipes, and other programs. Can be used as a graphics server program on a web site. Free time-limited demo version available.

Numerous statistical packages from companies acquired by SPSS Corp. Most of these demonstration versions expire after 30 days, and some have other limitations. Available products include:

- allCLEAR versions 3.5 and 4.5 (PC)
- GOLDMineR (PC)
- DeltaGraph (Macintosh)
- LogXact 2.1 (PC)
- PeakFit 4.06 (PC)
- QI Analyst 3.5 (PC)
- Remark Office OMR 3.0 (PC)
- SamplePower 1.2 (PC)
- SigmaGel 1.0 (PC)
- SPSS Diamond (PC)
- SigmaPlot 4.0 (PC)
- SigmaScan Pro 4.0 (PC)
- SigmaStat 2.0 (PC)
- SmartViewer (PC)

with common but unknown variance, and simple linear regression), examination of marginal distributions for arbitrary linear combinations of the location parameters, and the generation of predictive distributions.

IND -- Creation and manipulation of decision trees from data. For supervised classification and prediction in artificial intelligence and statistical pattern recognition. A tree is "grown" from data using a recursive partitioning algorithm to create a tree which (hopefully) has good prediction of classes on new data. IND improves on standard algorithms and introduces Bayesian and MML methods, producing more accurate class probability estimates that are important in applications like diagnosis. For UNIX systems. Currently available only in beta-test mode, and only to US citizens.

MANET -- ("Missings Are Now Equally Treated") Macintosh software for interactive graphics tools for data sets with missing values. Generates missing values chart, histograms & barcharts, boxplots & dotplots, scatterplots, mosaic plots, polygon plots, highlighted boxplots, interactive trellis displays, traces, context-sensitive interrogation, cues, redframing, selection sequences.

- StatXact 3.1 (PC)
- SYSTAT 7.0 (PC)
- TableCurve 2D 4.07 (PC)
- TableCurve 3D 3.01 (PC)
- WesVar Complex Samples (PC)

A large number of software demos are available for downloading from the website of SciencePlus, a distributor of scientific and related software (both full commercial packages and specialist academic tools). The list includes: ACTIV STATS, AGREE, AMOS, AQUAD, BIOFEEDBACK, BOJA, CADEMO, CART, CONTEST, CORWIN, DATADESK, DATA ENGINE, DBMS/Copy, EASYPLOT, EDWIN, ELI, E_PRIME, EQS, EQUITY, EQUIVTEST, ERTS, ERTSLAB, EXAMINER, EXPERT CHOICE, FASTTEST, GB-STAT, GETAREF, GLIMMIX, GOMAP, HIVIEW, HLM, ITEMAN, KWALITAN, LISREL, LPCM, MAPLE V, MELLAB, MEL 2, MICROCAT, MINITAB, MUDFOLD, NQUERY ADV., NSDSTAT+, OBSEVER, PARELLA, PEAKFIT, PLCA, POLYANALYST, RASCAL, REHACOM, SCRUTINY, SIGMA PLOT, SIGMA SCAN PRO, SIGMA STAT, SOLAS, STATISTICA, STRAD, STREAMS, SUPERLAB PRO, SUPERLAB LT, SYSTAT, TABLECURVE 2D/3D, TEXTANALYST, T-RASCH, TRIQ, UNISTAT, Vienna Test System, WINMIRA, WINROSA, XCALIBRE

GrafProg -- a Windows graphing program design, copy and save graphs generated by functions or by spreadsheet; also includes some statistical graphing processes.

StudioLine Photo Basic -- Photo editing software from H&M Software. Add descriptions to images, re-size photos for efficient e-mail transmission, print high-quality copies, display slide-shows, publish web-galleries, safe-keep images on CD or DVD. Version 2.2 has a new user interface, dual-monitor support, increased speed and other technical improvements. SmartUpdate feature checks for new versions. Has a web-board for user-to-user help.

WAFO -- Wave Analysis for Fatigue and Oceanography. A toolbox of Matlab (ver. 5.x / 6.x, for Windows & Unix) routines for statistical analysis and simulation of random waves and random loads. Tools are provided for analysis of measured data with routines for estimation of parameters in statistical distributions, estimation of spectra, plotting in probability papers, etc. Has routines for theoretical distributions of characteristic wave parameters from observed or theoretical power spectra of the sea. Another part is related to statistical

analysis of fatigue. The theoretical density of rainflow cycles can be computed from parameters of random loads. Has routines is included for modelling of switching loads (hidden Markov models). Also contains general statistical tools.	

Other Links to Collections of Free Software:

- Gene Shackman's page of links to free software packages. Contains sections for Statistical software, CDC/Census Bureau software, R, Other software, Lists of free stat software, Statistics with Excel, Mapping/GIS software, Non-statistical (but still useful) software, Office Suites (word processors -- stand-alone or web-based), Spreadsheets, Databases, Graphics, Web browsers / FTP clients, SUrvey software, Security software, and Miscellaneous.
- Citizendium's online article about free statistical software -- lots of links to free packages, but also other material about free stats software -- a brief history, reviews, advice about using the packages, and limitations of the packages.
- Links to Econometric Software (and **lots** of other general packages), maintained by *The Econometrics Journal*
- StatLib -- Software Archive, including Fortran source listings of hundreds of statistical and mathematical algorithms.

Return to John C. Pezzullo's Home Page, or send e-mail to jcp12345@gmail.com