Answers to Commonly Asked Statistics Questions

Mike Stockstill, Technical Support Statistician
SAS Institute Inc.

The purpose of this paper is to address some common 'how to' and 'what does this mean' users' questions collected by the statistics group in Technical Support. Product discussion includes SAS/IML®, SAS/INSIGHT®, SAS/STAT®, SAS/OR®, SAS/QC®, SAS/ETS®, and base SAS® software. The emphasis is on what can be done with the current releases rather than with future releases, although comments are provided where appropriate. Understanding the presentation requires little or no technical or statistical expertise. The paper should be most useful to site representatives who face a lot of stat-oriented questions. Topics addressed are organized by product as much as possible, and a miscellaneous section follows at the end.

The paper itself is based on our internal notes to each other and is not really intended to stand alone as any sort of documentation. We all contribute, and information is constantly updated. We use this information to help resolve as many calls as possible on initial contact and hope that you can share this with your fellow users.

Please note that this is not intended to be a complete description of any product, procedure, or macro. It does not replace any documentation. Instead, browse over it, and use it as a reference tool.

The best way to make use of this is by storing it as an ASCII file on your computer somewhere, and then use your 'find' or 'search' commands to look for text of interest. That is how we use it, and it works well. You may obtain an ASCII version of this paper in one of two ways (note that there are no plans to update it):

1. via anonymous ftp:
   `/techsup/download/samples/stat_and_iml/wusstext.sas`
2. via SIBBS: download the stat file named `wusstext.sas`

Occasionally, the answer to a question might make reference to a TS-###. This refers to a Technical Support Note or TS note. These are not available as a collection; however, we do provide them individually free of charge on an as-needed basis. Please do not request one unless you have a specific need — that way we can can continue to provide them without charge.

Other contributors to this paper include: David Schlotzhauer (who originated the file), Eddie Routten, Liz Shaw, Duane Hayes, Phil Gibbs, Annette Sanders, Annie Dudley, Cathy Maahs-Fladung, Donna Woodward, and Kathleen Kiernan.

SAS/ETS SOFTWARE:

PROC ARIMA

Q: Is there any additional documentation?
A: The *SAS/ETS Software: Applications Guide 1* provides examples of fitting univariate Box-Jenkins models in PROC ARIMA. Examples of estimating and forecasting transfer function and intervention models are also included. For a more theoretical reference, see the *SAS System for Forecasting Time Series*.

Q: What do I need to do if I get the warning `more values of input variable are needed`?
A: You get this message when you are forecasting either a regression model or a transfer function/intervention model and you do not supply future values of your input variable(s) for your LEAD= observations. You can supply future values by either concatenating them to the bottom of your DATA= data set or by fitting a model to your input series with IDENTIFY and ESTIMATE statements. Note that for intervention models (models with 0,1 dummy variables) you must supply values for your dummy variables in the DATA= data set.

Q: Is there any way to get the parameter estimates in an output data set?
A: As of Release 6.07.03, there are two ESTIMATE statement options that allow you to output the parameter estimates. They contain the same information, however, the data sets are structured differently, so you can chose the one most appropriate for your application needs. The options are OUTEST= and OUTMODEL=. These are documented in either the second edition of the *SAS/ETS User's Guide* or in SAS Technical Report P-231.

Q: Is there any way to get the statistics of fit in an output data set?
A: As of Release 6.07.03, the OUTSTAT= option in the ESTIMATE statement allows you to output the statistics of fit such as AIC, SBC, SSE, and the log likelihood. This is documented in either the second edition of the *SAS/ETS User's Guide* or in SAS Technical Report P-231.

PROC AUTOREG

Q: Can I fit ARCH or GARCH models?
A: As of Release 6.08, the AUTOREG procedure may be used to fit the following variations of GARCH models: generalized ARCH (GARCH), integrated GARCH (IGARCH), exponential GARCH (EGARCH) and GARCH-in-mean (GARCH-M). This functionality is documented in the second edition of the *SAS/ETS User's Guide*.

Q: For GARCH models, what distributions are supported for the error term?
A: Currently, only the standard Normal distribution is supported; however, PROC AUTOREG will support the Student's t distribution in Release 6.11. (Suggestions for alternative distributions are welcomed.)

Q: Is it possible to fit a Cochrane-Orcutt model?
A: No. The Cochrane-Orcutt method is similar to the default Yule-Walker method for first-order autocorrelation, except that the Yule-Walker method retains information from the first observation. To fit this model, specify the NLAG=1 option in the MODEL statement.

PROC EXPAND

Q: Is there an easy way to compute a moving average of a variable?
A: As of Release 6.08, the EXPAND procedure supports a number of transformation operations. These include weighted and unweighted moving averages, leads, lags, differences, logit and inverse logit transformations, to name a few. For the complete list, see pp. 411-412 of the second edition of the *SAS/ETS User's Guide*.

FORECASTING MENU SYSTEM

1st available: Release 6.11 (experimental in Release 6.07.03 and higher)

Q: How do I start it up?
A:

1. Under **Assist**, select **Planning tools**, then **Forecasting**, or
2. type FORECAST on the command line, or
3. via pmenus:
   Globals->Invoke Application->Forecasting

Q: Where is it documented?
A: Because it is experimental software, it is currently undocumented. A copy of the preliminary documentation is available upon request through Technical Support.

PROC MODEL

Q: Is there any way to impose bounds on the parameter estimates in my model?
A: Currently, the MODEL procedure does not support a BOUNDS statement, although one is planned for a future release. Several strategies are provided in the documentation for imposing bounds on the parameters. These are described on pp. 589-590 of the second edition of the *SAS/ETS User's Guide*.

Q: Can I do Generalized Method of Moments Estimation?
A: Yes, as of Release 6.08, the MODEL procedure supports GMM estimation by specifying GMM on the FIT statement. The BARTLETT, PARZEN, and Quadratic Spectral kernels are supported for this method. For the details, see pp. 555-557 of the second edition of the *SAS/ETS User's Guide*.

Q: Can I do Full Information Maximum Likelihood estimation?
A: Yes, as of Release 6.08, the MODEL procedure supports FIML estimation by specifying FIML on the FIT statement. Several estimators of the variance covariance matrix of the parameter estimates are available. For the details, see pp. 558-560 of the second edition of the *SAS/ETS User's Guide*.

PROC TSCSREG

1st available: Release 6.07 (experimental 6.04, preliminary doc available on request through Tech Support: TS-245); doc: 2nd ed. *SAS/ETS User's Guide*

Q: Is TSCSREG available under Release 6.04?
A: An experimental version of the procedure is available upon request through Technical Support.

Q: Why do I get the warning `phi matrix is singular` when I run PARKS method?
A: Any time the number of cross sections is greater than the number of time series observations per cross section, PROC TSCSREG will produce the warning. (This is analogous to running seemingly unrelated regression with fewer observations than equations in the model.) The only circumvention is to reduce the number of cross sections.

Q: PROC TSCSREG requires the same number of time series observations in each cross section. Is there any way around this restriction? I don't want to omit any of my data from the analysis.
A: There is no way around this current limitation in TSCSREG. However, if you have access to SAS/STAT, then the MIXED procedure can be used to fit models that are very close to the Fuller-Battese and Da Silva methods in PROC TSCSREG. Following are examples of the corresponding specifications:

```
proc tscsreg data=a fuller;
 /* fuller-battese model in TSCSREG */
   model y=x1 x2 x3 / covb;
   id state year;
  run;

proc mixed data=a;
 /* fuller-battese model in MIXED   */
   class state year;
   model y=x1 x2 x3 / s p;
   random state year;
  run;

proc tscsreg data=a dasilva m=5;
 /* dasilva model in TSCSREG   */
   model y = x1 x2 x3;
   id state year;
  run;

proc mixed data=a;
 /* dasilva model in MIXED          */
 /* please note that the estimates  */
 /* will not be identical since     */
 /* TSCSREG uses Seely's method and */
 /* MIXED uses REML,                */
 /* however, they should be close   */
  class state year;
  model y = x1 x2 x3 / s;
  random state year;
  repeated / type=toep(6) sub=state;
 run;
```

MACROS

Q: Is there any way to compute the augmented Dickey-Fuller test for unit root?
A: Yes, as of Release 6.08, two macros included in the SAS macro autocall library may be used to compute the Dickey-Fuller test for unit root and its corresponding p-value. The macros are called %DFTEST and %DFPVALUE. The %DFTEST automatically calls the %DFPVALUE macro; however, the %DFPVALUE may also be run on its own to compute the p-values for tests statistics that have already been computed. For more information regarding these and several other macros, see pp. 945-957 of the second edition of the *SAS/ETS User's Guide*.

SAS/IML Software

Q: Can SAS/IML be run in batch mode?
A: Yes, it can, but it will still execute statement by statement.

Q: How large of a matrix can I define in my PROC IML job?
A: Below are the hard-coded matrix size limitations in SAS/IML software by platform and SAS release:

MS-DOS, PC-DOS	6.03, 6.04	4095 elements
UNIX	6.03	32767 rows x 32767 cols
OS/2	6.06	32767 rows x 32767 cols
Windows	6.08	32767 rows x 32767 cols
OS/2	6.08	limited only by memory
UNIX	6.07, 6.09	limited only by memory
Windows NT	6.09	limited only by memory
MVS, CMS, VMS	6.06 and higher	limited only by memory

*** Starting with Release 6.10, *ALL* SAS/IML releases will be unlimited (except by available memory).

TREEDISC MACRO

1st available: Release 6.10 sample library (doc.: internal; see header)

Q: What does it require?
A: SAS/IML. SAS/OR is required to draw the tree but the tree can be listed without SAS/OR. Release 6.08 or higher is recommended, but it definitely will not run on Release 6.04 or earlier.

Q: What does it do?
A: The %TREEDISC macro generates a SAS data set that describes a decision tree computed from an input data set to predict a specified categorical dependent variable from one or more other predictor variables. The tree can be listed or drawn or used to generate code for a SAS DATA step to classify observations.

Q: How do I get it?

A: You can get them via anonymous ftp:
`/techsup/download/samples/stat_and_iml/treedisc.sas`
`/techsup/download/samples/stat_and_iml/xmacro.sas`
SIBBS:
stat file `treedisc.sas`
stat file `xmacro.sas`

To use, see documentation in headers.

Q: What are the three sets of statistics printed at each node?
A: VAL: Stands for VALUES. The values you see are the values of the predictor variable by which the node was split. For the root (the trunk of the tree), the values of the dependent variable are given. Missing values for character variables are noted with an '*'. COU: Stands for COUNT. This category contains the number of observations in each category of the dependent variable on the node. PVAL: Stands for PVALUES. The values listed here are the two smallest adjusted p values that could be used for further splits. If there are no p values present, %TREEDISC will not split again at that branch at that node.

Q: Why weren't some of my observations classified?
A: Whenever an observation could be assigned to more than one node, it is considered *tied*. The %TREEDISC variable TIE_ shows the value two or more. The observation is dropped from the analysis.

SAS/INSIGHT Software

1st available: Release 6.07.01 (Release 6.07.01–.03 doc.: *SAS INSIGHT User's Guide*, 1st ed.; Release 6.08 and later doc.: 2nd ed.)

Q: How do you get parameter estimates for polynomial fit?
A: If you fit a model with X in it and then use the menu selection `CURVES . . . POLYNOMIAL . . . 2` (to get a quadratic fit, for instance), the curve is shown but the parameters are not. To get the parameters (Releases 6.08 and 6.09), select `ANALYZE . . . FIT(Y X)`. In the FIT window, select the Y and X variables as before, then select the X variable in the rightmost window and click on the EXPAND to add terms up to the order indicated by the value below EXPAND. You can change the EXPAND value. Then click on OK or RUN to fit the model. The parameters will be displayed. Starting in Release 6.10, there will be a `Parameter Estimates` check box in the `CURVES . . . POLYNOMIAL` dialog box that, if selected, will add a table of the parameter estimates for the requested model.

SAS/OR SOFTWARE:

PROC LP

Q: What size problem will PROC LP handle? How many variables and constraints can I have in my problem?
A: There is no definite answer to this question. The size of the problem is very data dependent. We have run 20,000 variables by 20,000 constraint problems successfully but have had difficulty solving much smaller problems. If there are integer or binary variables, the problem becomes much more difficult to solve. A large integer problem is on the order of 2000 variables.

PROC NLP

1st available: experimental Release 6.08; production Release 6.10

Q: Where is it documented?
A: PROC NLP is available as an EXPERIMENTAL procedure beginning in Release 6.08. The procedure solves nonlinear programming problems, that is, optimizing a nonlinear objective function when they have linear or boundary constraints. There is an extended users guide available upon request.

Q: Can PROC NLP handle nonlinear constraints?
A: No. But beginning in Release 6.09, there are routines in PROC IML that will. An internal draft document that describes these routines is available upon request. The IML routines are NLPNMS, NLPQUA, NLPCG, NLPQN, NLPDD, NLPTR, NLPNRA, NLPNRR, NLPLM, and NLPHQN. Each routine performs the optimization using a different algorithm.

Q: Does SAS have a function or procedure that will maximize a likelihood function?
A: Yes— PROC NLP will maximize a likelihood function as long as it can be expressed in DATA step code. There are examples available in the test stream, though PROC NLP is not difficult to use.

SAS/QC SOFTWARE:

ADX MENU SYSTEM

Q: Where is it documented?
A: Official documentation is now available. You can request "ADX Menu System for Design of Experiments" from Book Sales (919-677-8000, x7001). Or, you may have a free copy of the old TS-212, which is a yellow booklet titled "ADX Menu System Examples".

Q: What is needed to use it?
A: SAS/QC and SAS/STAT are required. SAS/GRAPH® software is necessary to produce high-resolution graphs. SAS/FSP® software is necessary to be able to add response information from within ADX (otherwise, you will need to write and submit a small DATA step program to add response data). Consequently, SAS/FSP is highly desirable.

Q: If a design has already been created, how can it be imported for analysis?
A: If the design is one of the types that ADX can create, then you can import it by clicking on `File: New design: Add existing data set` and filling in the resulting window.

ADX MACROS

1st available: Release 6.03 doc.: *SAS/QC Software Reference, Version 6.*

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

PROC CAPABILITY

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

Q: Is there any additional documentation?
A: Yes. TS-291: "Process Capability Analysis Using SAS/QC Software, Release 6.08" (the orange booklet) contains documentation for new Release 6.08 features including the INSET and COMPHISTOGRAM statements. We can mail it upon request.

Q: Can I overlay more than one curve from the same family of distributions onto a histogram?
A: No, not in PROC CAPABILITY. You can use PROC GREPLAY to replay multiple plots on top of each other.

PROC CUSUM

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

PROC FACTEX

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

Q: How can I completely randomize the output plan?
A: Use the RANDOMIZE option on the OUTPUT statement. However, if you are also using either the DESIGNREP= or POINTREP= options, then RANDOMIZE won't do it. After you output the plan, add a random variable and then sort on it, like this:

```
data outplan; set outplan;
  random=ranuni(123);
  run;
proc sort; by random; run;
```

PROC MACONTROL

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

PROC OPTEX

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

Q: How can I completely randomize the output plan?
A: You can't with PROC OPTEX. After you output the plan, add a random variable and then sort on it, like this:

```
data outplan; set outplan;
  random=ranuni(123);
  run;
proc sort; by random; run;
```

PROC SHEWHART

Q: *SAS/QC Software: Reference* doesn't show examples. Are there any?
A: Yes. Examples of all SAS/QC procedures are in SAS Technical Report P-188.

Q: Any additional documentation?
A: Yes. TS-175, "Selected SAS/QC Software Examples, Release 6.06" (light blue booklet) is a companion to Technical Report P-188. It contains additional documentation and examples for PROC SHEWHART, including star charts, notched box charts, and box charts with varying widths. We can mail it upon request.

Q: The standard deviation computed by PROC SHEWHART is different from the one computed by PROC MEANS/UNIVARIATE. Why?
A: The formulas used to compute variability are different. PROC MEANS and PROC UNIVARIATE use the usual 'sum of squared deviations from the mean' formula, whereas in control charting, we compute an unweighted average of subgroup standard deviations or ranges. See "Computational Details" in *SAS/QC Software: Reference* for the formulas used in quality control.

Q: Limits from a LIMITS= data set are ignored, not used. Why?
A: By default, control limits will be computed based on the input DATA= data set. If you want to use the limits from a LIMITS= data set, you must also specify the READLIMITS option on the chart statement. The reason you must tell PROC SHEWHART to use the limits is so that you can have multiple chart statements in one run of the procedure, and you choose which charts will use previously determined limits and which charts will have limits based on the raw input data.

Q: How do I control the number of pages PROC SHEWHART uses to print a chart, or how can I put more subgroups on one page?
A: Use the NPANELPOS= option on any chart statement to tell PROC SHEWHART how many subgroup positions to put on a page. For example, if you have 100 subgroups and you want your control chart to come out on one page, use NPANELPOS=100 in the chart statement. If you want this same control chart to come out on two pages, use NPANELPOS=50. The default is NPANELPOS=50 for all charts except boxchart, where the default is 20.

SQC MENU SYSTEM

1st available: Release 6.03 (doc.: *SAS/QC Software: SQC Menu System, Version 6, First Edition* and on-line tutorial)

Q: How do I start it up?
A: For Releases 6.03 and 6.04: Type the following on the command line: `af c=sashelp.sqc.qcmenu.program`. Releases 6.06, 6.07.01, 6.07.02:

1. Under **Assist**, select **Planning Tools**, then **Quality Cntl**, or
2. type SQC on the command line (Release 6.07.03 and later):
   a. Under **Assist**, select **Data Analysis**, then **Quality Cntl**, or

b. type SQC on the command line, or
c. via pmenus:
```
Globals->Invoke Application->Data Analysis->
Quality Control
```

SAS/STAT SOFTWARE—ANOVA PROCEDURES:

Q: How can I get multiple comparisons on interaction means?
A: See the SUGI 18 paper "Strategies for Performing Multiple Comparisons on Means" by Jenny Kendall. It is also available as TS-282 (ask for TS-116 too).

PROC GENMOD

1st available: Release 6.09 (experimental in Release 6.08; production in Release 6.08 4th maintenance = 6.08 TS415); doc.: Technical Report P-243

Q: What are some good references?
A: The McCullagh and Nelder reference at the end of the GENMOD chapter is very good and has many examples. The Dobson reference provides a good, low-level, introductory reference.

Q: Does it do GAMs (Generalized Additive Models) or GEEs (Generalized Estimating Equations)?
A: No. However, there are some user-written macros available.

Q: Can it fit data using a negative binomial distribution?
A: Yes. The file GENMODX3.SAS in the Release 6.10 sample library shows an example. It is available upon request.

PROC GLM

Q: Are there any additional references and documentation?
A: *SAS System for Linear Models, Third Edition* also the "DETAILS" section and the introductory chapters in the *SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 and 2*.

Q: How can I write parameter estimates to an output data set?
A: In Release 6.07.03 or later, use PROC MIXED with a MAKE statement selecting the 'SolutionF' table for output. For example, you can output the parameters of the following GLM analysis with MIXED:

```
proc glm;   class group; model y=group x*group
            / noint solution;
proc mixed; class group; model y=group x*group
            / noint solution;
            make 'SolutionF' out=parmests;
```

Prior to Release 6.07.03, the options you have are

1. PROC PRINTTO; or
2. Coding your own dummy variables and using PROC REG with the OUTEST= option. Which option is better depends on the user, the number of CLASS variables and levels, and the complexity of the model.

Q: How can I obtain multiple comparisons on interaction means?
A: See the SUGI 18 paper "Strategies for Performing Multiple Comparisons on Means" by Jenny Kendall. It is also available as TS-282 (ask for TS-116 too).

Q: How can I obtain multiple comparisons for my repeated measures?
A: See the SUGI 15 paper "Performance of Multiple Comparisons in Repeated Measures Designs under Nonsphericity" by Steve Thompson. It is also available as TS-235.

Q: My hand computations for the expected mean squares do not match the output. Why is this?
A: See pp. 197-198 of the *SAS System for Linear Models*. There is disagreement as to the best method for computing expected mean squares. These pages from SSLM explain our reasoning to a limited extent and give the relevant references.

Q: What does a non-estimable CONTRAST (or ESTIMATE) mean?
A: See TS-116 or *SAS System for Linear Models*. The most common error encountered here is not assigning 'main effect' coefficients when they are involved in the comparison.

Q: What do I do if I get an insufficient memory error or some kind of system ABEND while running PROC GLM?
A: See SAS Usage Notes 5567 and 6177. You need to compute how much memory the model needs, then add about 2.5M for GLM overhead.

Q: When should I use PROC MIXED instead of PROC GLM?
A: Anytime you have:

1. a RANDOM statement or
2. a TEST statement or
3. a nested term that is likely a random term in the MODEL statement.
4. an E= option on LSMEANS, MEANS, or CONTRAST statement.
5. SOLUTION option in the MODEL statement output that you want in a data set
6. ESTIMATE statement output that you want in a data set.
7. a repeated measures problem with a covariate or missing values on the subjects. To get started with PROC MIXED, read the MIXED sections in Technical Reports P-229 and P-242 (Release 6.08). In general, all RANDOM effects will go in the RANDOM statement and **NOT** in the MODEL statement in PROC MIXED.

PROC MIXED

1st available: Release 6.07.03 (doc.: Technical Report P-229)

Q: Is there any additional reference documentation?
A: TS-260: A Tutorial on Mixed Models. SAS Technical Report 229 is the primary documentation; SAS Technical Report 242 contains changes and enhancements. A general reference on mixed models would be the *Southern Cooperative Series Bulletin 343, Applications of Mixed Models in Agriculture and Related Disciplines*, available from the Louisiana State University Experiment Station in Baton Rouge.

Q: Can PROC MIXED reproduce a PROC GLM analysis?
A: Whether you will be able to do this is highly dependent on the model. As a general rule, if you have a RANDOM or REPEATED statement, use PROC MIXED instead of PROC GLM.

Q: How can I get the sums of squares for my fixed effects?
A: You can't. PROC MIXED uses maximum likelihood methods and does not compute sums of squares. See CONTRAST statement documentation for an explanation of how the *F* statistic is constructed.

Q: How can I get my parameter estimates?
A: For fixed effects, use the SOLUTION option in the MODEL statement. For random effects, use the SOLUTION option in the RANDOM statement.

Q: PROC MIXED is taking a very long time or a lot of computer resources or is causing an ABEND of some kind. Is there something I can do?
A: See SAS Usage Notes 5568 and 6176. The MIXED procedure is potentially one the biggest resource hogs in SAS/STAT.

Q: Where is my PROC MIXED output? Or, why is my PROC MIXED output going directly to the printer?
A: See SAS Usage Note 7068. This is an output delivery system issue.

PROC MULTTEST

1st available: Release 6.07.03 (doc.: Technical Report P-229)

Q: How does one compute a trend test for proportions?
A: For example, if Y indicates response or no response and X is the dose amount of a drug, one can test that the proportion of response increases with dose using the Cochran-Armitage test in PROC MULTTEST (with CLASS X; TEST CA(Y);). This can also be done in PROC LOGISTIC (with MODEL Y=X;) by looking at the score test printed in the 'Criteria for Assessing Model Fit' table.

PROC NPAR1WAY

Q: Can I get nonparametric multiple comparisons from PROC NPAR1WAY?
A: No. It is on the ballot. Some users run PROC GLM on the ranks of the data. Ron Iman has written some books on this.

PROC TTEST

Q: Can PROC TTEST test the difference between two groups using summary statistics?
A: No, but it can be done easily in a DATA step using the formulas in the PROC TTEST documentation, along with the PROBF and PROBT functions.

PROC VARCOMP

Q: PROC VARCOMP and PROC GLM results don't agree — why?
A: They will agree if the model is specified in such a way that they are fitting exactly the same Type I model in both procedures. The order of the effects in the MODEL statements is important, as well as appropriate specification of what is a FIXED effect in PROC VARCOMP, and what is a RANDOM effect in PROC GLM.

Q: How can I write the results of my PROC VARCOMP output to a data set?
A: You can't. Try PROC MIXED.

Q: How does PROC VARCOMP compare to PROC MIXED?
A: PROC MIXED does not perform a Type I analysis. For certain models, PROC MIXED and VARCOMP will not agree, and that is due to the nature of the algorithms used. MIXED and VARCOMP come up with starting values differently. As a general rule, MIXED is a more robust algorithm, but sometimes people are happier with VARCOMP results.

SAS/STAT SOFTWARE—CATEGORICAL PROCEDURES:

PROC CATMOD

Q: Is there any additional documentation?
A: There is a book about using SAS for categorical analysis underway. You may purchase the Categorical Data Analysis Course Notes from Book Sales. This book is not designed to be used by itself, but it is pretty well done and may still be useful on its own.

Q: Are there any recommended references?
A: Agresti (1990), *Categorical Data Analysis*, published by John Wiley & Sons, Inc., provides very good survey and talks some about SAS. Freeman (1987), *Applied Categorical Data Analysis*, is also very good and is perhaps a bit lower-level than Agresti. He also shows some SAS code. There are several Sage publications that may be instructional.

Q: What should I do if the procedure is running out of memory or taking a lot of time?
A: As the number of parameters in the model increases, so does the amount of memory and time needed. The number of parameters in the model is a function of the number of variables and the number of levels in each variable. Generally, PROC CATMOD is not intended to handle ANY variables with many levels.

Q: I think I am getting backward parameter estimate signs — why? How can I fix it?
A: PROC CATMOD models the probability of the LOWER response level. For example, if no RESPONSE statement is specified and if the response Y is binary with levels 0 and 1, CATMOD models a quantity (called a logit) that increases and decreases with Pr(Y=0). This causes the parameter estimate signs to be reversed from modeling Pr(Y=1). If you want to model Pr(Y=1), then do this:

```
proc sort; by descending y;
proc catmod order=data ....
```

Q: Which parameter estimate goes with which independent variable level? Why aren't there as many parameters as levels of my independent variable?
A: See Example 1 in the *SAS/STAT User's Guide, Volume 1*. Variable level ordering is given by the default POPULATION PROFILE section or by the ONEWAY option (both shown on page 479). Parameters are on top of page 481. Parameter 2 is the effect of the first level of TRT (a), parm 3 is the effect of the 2nd level of TRT (b), parm 4 is the effect of the 3rd level of TRT (c). Notice that there is no parm for the last level of TRT (d). That is because of the sum-to-zero constraint of the model. The effect of the last level is constrained to be the negative of the sum of the other parameters (−(parm1+parm2+parm3)).

Q: What do I do about infinite parameters or redundant effects messages?
A: These appear when some parameters of the model are detected to be infinite (see middle of page 470). Parameters can become infinite when there are more parameters in the model than the data can support or when some variable(s) perfectly predict the response. The only thing to do is reduce the number of effects in the MODEL statement and/or the number of levels per variable (both of which will reduce the number of parameters in the model). If any variable in the MODEL statement has many levels (variables in the DIRECT statement, if used, are likely candidates here), try categorizing it to reduce its number of levels.

Q: What does this error message mean and what can I do to fix it: `The response functions are linearly dependent since the number of functions per population, x, is greater than or equal to the number of response levels, y, in population z (where x, y and z are integers).`
A: It means the data are too sparse for the model specified. One way to fix this is to decrease the number of populations. The number of populations is a function of the number of independent variables appearing on the MODEL statement AND the number of distinct values appearing in each of these variables. The number

of populations can be decreased by either dropping independent variables or merging (also called *collapsing*) distinct values of one or more independent variable. The obvious place to start is to collapse the values of the variable with the most values. Using the FREQ option in the MODEL statement will print the underlying two-way table that CATMOD is trying to analyze. Each row is a population, and each column is a response level. This problem occurs because of too many zeros appearing in the table. The distribution of zeros in the FREQ table should suggest the best way to collapse or drop variables. Note that another way to reduce the number of zeros in the table would be to reduce the number of columns by collapsing the response variable values.

PROC FREQ

Q: How do I input cell counts instead of raw data?
A: Use the WEIGHT statement to specify a variable containing cell counts. See any of the examples at the end of the PROC FREQ chapter in the *SAS/STAT User's Guide*.

Q: How do I get a chi-square test on a one-way (1xC or Rx1) table? Or, how do I compute a goodness of fit test?
A: PROC FREQ does not do any tests on one-way tables. If you want to test that the cell frequencies in a one-way table are equal or follow some distribution that you can specify, see "Computing Goodness-of-fit Chi-square Statistics with SAS Software" by David Schlotzhauer, *Observations*®, Fourth Quarter 1991. It is also available as TS-176. Also, PEARSON.SAS in the sample library will give chi-square tests of equal cell frequencies.

Q: How do I compute agreement statistics such as McNemar's test, Cochran's Q, or kappa statistics?
A: Page 129 of the *SAS/STAT User's Guide* shows how to compute McNemar's test. The same technique is used to get Cochran's Q when the table is larger than 2x2. Kappa statistics can be computed with PROC CATMOD, but it is not easy. A complex RESPONSE statement must be written for each situation. See the 1987 SUGI paper "Generating Kappa Statistics and Testing Useful Hypotheses with PROC CATMOD," page 1149, by Robert Terry. The paper is also available as TS-118. In Release 6.10 an option called AGREE will be added to the TABLES statement in PROC FREQ that will make it easy to get kappa statistics. It will also simplify computation of McNemar's test and Cochran's Q statistic.

Q: How do I do multiple comparisons or tests on subtables if I get a significant Pearson chi-square?
A: There are no multiple comparison methods per se. However, you could use the CONTRAST statement in PROC CATMOD to test for row (or column) differences. Also, correspondence analysis is often used as a way to visualize the non-independence in a table. See the examples in the PROC CORRESP chapter of the *SAS/STAT User's Guide*. Or, you can just do additional Pearson chi-square tests on subtables. The WHERE statement is helpful for selecting subsets of rows or columns. Define formats to merge rows or columns.

PROC LOGISTIC

1st available: Release 6.04 (Release 6.04 doc.: SAS Technical Report P-200)

Q: Are there any recommended references or additional documentation?
A: The Hosmer and Lemeshow reference at the end of the PROC LOGISTIC chapter is probably the best and is fairly readable. We are producing a book on using SAS for categorical data analysis.

Q: What do I do if I get backward parameter estimate signs — why? How can I fix it?
A: PROC LOGISTIC models the probability of the LOWER response level. For example, for a binary response, Y, with levels 0 and 1, LOGISTIC models a quantity (called a logit) that increases and decreases with Pr(Y=0). This causes the parameter estimate signs to be reversed from modeling Pr(Y=1). If you want to model Pr(Y=1), then use the DESCENDING option (6.07.03 and later), or do:

```
proc sort; by descending y;
proc logistic order=data ....
```

If you need more information see "Some Issues in Using the LOGISTIC Procedure for Binary Logistic Regression" by David Schlotzhauer, *Observations*, Third Quarter 1993. It is also available as TS-274.

Q: Odds ratios — how do I get them? How are they computed/interpreted?
A: See "Some Issues in Using the LOGISTIC Procedure for Binary Logistic Regression" by David Schlotzhauer, *Observations*, Third Quarter 1993. It is also available as TS-274.

Q: Predicted probabilities — how are they computed/interpreted?
A: See "Some Issues in Using PROC LOGISTIC for Binary Logistic Regression" by David Schlotzhauer in *Observations* Third Quarter 1993. It is also available as TS-274.

Q: How do I classify/score new observations?
A: Add the new observations to the original data set, setting the response to missing for all of the new observations. PROC LOGISTIC ignores all observations with missing response values (so the fit will be identical to using just the original data set). Then use the OUTPUT statement to request the OUT= data set and predicted probabilities (P=*varname*). The OUT= data set will contain predicted probabilities for all observations, including the new ones. PROC SCORE will **NOT** do it. See TS-274.

Q: What is a multinomial logit model (or discrete choice or McFadden's conditional logistic model) and how do I fit it?
A: These are models for data in which the response is usually a set of choices (and, therefore, is nominally-scaled) and at least some of the independent variables indicate characteristics of the choices (cost, size, attractiveness) instead of characteristics of the subject or chooser as is usually the case. The effect of an independent variable becomes conditional on the subject's choice between two alternatives and depends on the distance between these alternatives' values on the variable. Such models can ONLY be fit with PROC PHREG. See "Multinomial Logit, Discrete Choice Modeling" by Warren Kuhfeld. It is available as TS-273. Note that in models that can be fit with PROC CATMOD and PROC LOGISTIC, the independent variables are always characteristics of the chooser, not of the choices. In this case, PROC CATMOD can be used for a nominally- or ordinally-scaled response. PROC LOGISTIC can be used for an ordinally-scaled response and is preferable to PROC CATMOD if any of the independent variables are continuous.

Q: Non-convergence — why is it happening? How can I fix it?
A: This is a common problem. It is caused by some parameters of the model becoming theoretically infinite. This can happen when the model perfectly predicts the response. PROC LOGISTIC does not converge unless the largest change among the parameters is small, and this will never happen when a parameter is infinite. The only way to fix this is to reduce the number of variables and/or change continuous variables to categorical. There is no way to know EXACTLY which variable to eliminate/categorize — it's a trial and error sort of thing. This problem typically happens with sparse data, where *sparse* means that there are few or no repeats of each setting of the covariates. Or, instead of reducing or

altering the variables, you could use an alternative rule to decide when to stop the iterations. A way to do this is presented in the SUGI 18 Proceedings, "A Tutorial on Logistic Regression" by Ying So.

Q: WEIGHT and FREQ: what's the difference? Why do all of my effects become significant if I include a WEIGHT statement?
A: A WEIGHT statement should be used when you want to unequally weight the contributions of observations to the likelihood function, which is maximized by the procedure. Weights are often used to adjust for the use of stratified, rather than random, sampling in a study (see below). The WEIGHT statement does not change the value of N (sample size) used in formulas for some statistics (listed in the WEIGHT statement documentation). The FREQ statement should be used to enter summarized data in which only the unique combinations of independent variable settings are stored along with a variable indicating the number of times each occurred in the raw data (exactly like entering just the cell counts of a contingency table as opposed to raw data). The values **MUST** be integers; weights may be fractional. The FREQ statement **DOES** affect N. Using a given integer-valued variable as a WEIGHT variable or as a FREQ variable results in the same parameter estimates and standard errors. However, inflating the values of a FREQ or WEIGHT variable (such as multiplying them by a constant) will drive all effects toward significance. For this reason, weights are typically 'normalized' so that they sum to the actual sample size. You can request that the procedure normalize your weights by using the undocumented NORMWT option in the PROC statement.

Q: If I draw samples from each response level (that is, stratified sampling as in a retrospective study), do I have to weight my data to make the results meaningful?
A: The odds ratio and its confidence limits are unaffected by the sampling scheme, so no weighting is needed to report them. Except for the intercept, the parameter estimates and their standard errors are also ok. However, because the intercept is affected, any computation based on the full set of parameters is incorrect, such as the predicted event probabilities, differences or ratios of event probabilities (the ratio is called the relative risk), and false positive and negative rates. If the event probabilities are very small in each group to be compared (no, we can't say how small is small), then the odds ratio provides an estimate of the relative risk. See Agresti's book, *Categorical Data Analysis*, pp. 8-18, for more information on the relative risk. For more information on handling data from retrospective studies, see Chapter 6 of the Hosmer and Lemeshow reference or Chapter 7 of D. Collett's book *Modelling Binary Data* (1991), Chapman & Hall.

PROC PROBIT

1st available: Release 6.03 update (doc.: SAS Technical Report P-179)

Q: Is it available on the PC under Releases 6.03 or 6.04?
A: It was not available when Release 6.03 was released but was when the Release 6.03 update was shipped. It was documented in SAS Technical Report P-179. It has been in all releases since (including Release 6.04).

Q: Why do I get backward parameter estimate signs? How can I fix it?
A: PROC PROBIT models the probability of the LOWER response level. For example, for a binary response, Y, with levels 0 and 1, PROBIT models a quantity that increases and decreases with Pr(Y=0). This causes the parameter estimate signs to be reversed from modeling Pr(Y=1). If you want to model Pr(Y=1), then do:

```
proc sort; by descending y;
proc probit order=data ....
```

Q: How do I classify/score new observations?
A: Add the new observations to the original data set, setting the response to missing for all of the new observations. PROC LOGISTIC ignores all observations with missing response values (so the fit will be identical to using just the original data set). Then use the OUTPUT statement to request the OUT= data set and predicted probabilities (P=*varname*). The OUT= data set will contain predicted probabilities for all observations, including the new ones. PROC SCORE will **NOT** do it!

Q: Why are some of (or all) the inverse confidence limits missing?
A: If the first parameter estimate (after INTERCPT) is not strongly significant, this often happens and is correct. Keep in mind that you're trying to get a confidence interval on the INDEPENDENT variable that would generate a given response. Think of a plot of response on the vertical axis and the first independent variable on the horizontal axis (any other independents are held fixed). Imagine a regression line with confidence bands around it. Now draw a horizontal line at the response of interest. Where the horizontal line hits the confidence bands defines the inverse confidence interval on the independent variable for that response. Now, if there is a strongly significant slope, the horizontal line should intersect both confidence bands. But, if the slope is near zero (the regression line is flat), then the horizontal line may contact only one or neither band. When this happens, the confidence interval is just an upper or lower bound or is unbounded.

Q: Why can't I get the OUTEST= data set?
A: The OUTEST= data set cannot be generated if the CLASS statement is used. See SAS Usage Note 6000 for a possible work-around.

Q: Can it compute inverse Mill's ratio, or can it do Heckman's two-stage model to correct for sample-selection?
A: No, but there is a user-contributed program that does both of these. Via anonymous ftp:
`get /techsup/download/samples/stat_and_iml/heckman.sas`
SIBBS: stat file `heckman.sas`

Q: Can you give more details about LACKFIT tests or heterogeneity factor with respect to sample size or degrees of freedom?
A: Because continuous variables may appear in the model, the lack-of-fit statistics and the heterogeneity factor are not computed within replicate groups and then pooled. However, the data should be presorted by the covariates (that is, independent variables) to correctly compute the statistics. Further, the LACKFIT tests are not good tests of lack-of-fit if the number of replicates in some settings of the covariates is small. The heterogeneity factor is an estimate of overdispersion and is simply the Pearson chi-square statistic divided by it's degrees of freedom. This is appropriate when there is little or no replication. We intend to add a "replicates analysis" option in Version 7 that will compute pooled, within-replicate-group LACKFIT tests and overdispersion. A good discussion of these issues is given in McCullagh and Nelder, *Generalized Linear Models, 2nd edition*, (1984, pp. 118-122 and 126-128).

SAS/STAT SOFTWARE—CLUSTERING PROCEDURES:

PROC CLUSTER

Q: How can I tell how many clusters are in my data set?
A: The best place to look for information regarding the number of clusters is in the "Introduction to Clustering Procedures" chapter in the *SAS/STAT User's Guide*. See also the papers by Milligan & Cooper and Cooper & Milligan referenced there. You can examine several statistics in the printed output to determine the number of clusters if you specify the options PSEUDO and CCC. To interpret the

- pseudo T statistic: start at the top of the printed output, and look for the first relatively large value, then move back up one cluster.
- pseudo F: look for a relatively large value.
- CCC: values greater than 2 or 3 indicate good clusters; values between 0 and 2 indicate potential clusters (but should be taken with caution); large negative values may indicate outliers.
- R-Square: look for a value that explains as much variance as you think appropriate. Milligan and Cooper demonstrated that changes in the R-square are not very useful for estimating the number of clusters, but it may be useful if you are interested solely in data reduction.

Q: How can I tell which observations go in which cluster?
A: First, run PROC CLUSTER and create an OUTTREE= data set. Once you have decided how many clusters you want, run PROC TREE on the OUTTREE= data set, specify N= the number of clusters, and OUT= the name of a data set to create. The created data set will have a new variable named CLUSTER, the values of which indicate the cluster member for each observation.

Q: Can I cluster binary data?
A: PROC CLUSTER expects all raw data to be continuous. If you want to cluster binary data, create a distance matrix (you may be able to use the %DISTANCE macro to do this for Release 6.08 and later; Release 6.07 and later on MVS, CMS, and VMS) as you think it should be defined, and then use this as input to PROC CLUSTER. PROC FASTCLUS will not accept distance data.

Q: How many observations can I cluster?
A: PROC CLUSTER is not practical for very large data sets. See example 3 in the PROC CLUSTER chapter — about half way through the example, it shows how to form preliminary clusters with PROC FASTCLUS, cluster these preliminary clusters with CLUSTER, and then reassign the original observations. How BIG a data set has to be before it is considered BIG depends on how much memory your machine has, how long you are willing to wait, and how much money you are willing to spend on CPU.

Q: I've already clustered one set of data, and I want to take a new data set and put the observations in the same clusters. How do I do this?
A: There are two ways. One uses PROC FASTCLUS; the other uses PROC DISCRIM.

1. PROC FASTCLUS: First, compute the means from each of your current clusters to use as a SEED= data set in PROC FASTCLUS. Then run PROC FASTCLUS with DATA=*your new data set*, SEED=*the data set of means from your current clusters*, MAXITER=0, REPLACE=NONE, and MAXC=*the number of clusters you have*. Be sure to use a VAR statement listing the same variables as you used in PROC CLUSTER.
2. PROC DISCRIM: Use your data set of cluster-identified observations as the DATA= data set and your data set of new observations as the TESTDATA= data set. List the cluster-identifying variable in the CLASS statement. Use the same VAR statement as you used in CLUSTER. See example 5 in the DISCRIM chapter of the *SAS/STAT User's Guide, Volume 1* (if Release 6.06 or Release 6.03).

PROC TREE

Q: Can I get high resolution dendrograms (tree diagrams)?
A: PROC TREE does not currently produce high resolution trees, but it will beginning with Release 6.11. Currently, there is a user-written macro that will work, scheduled for publication in *Observations*, First Quarter 1995. The code is available upon request.

SAS/STAT SOFTWARE—DISCRIMINANT PROCEDURES:

PROC CANDISC

Q: Which variables are most important (are best discriminators)?
A: PROC CANDISC (or PROC DISCRIM with the CANONICAL option) can tell you how important each variable is via the structure coefficients. The structure coefficients are just the correlations between the original variables and the canonical variables (what some people call loadings). The discriminant function coefficients from DISCRIM are **NOT** the best way to determine variable importance because correlation among the variables can make their interpretation in that way misleading.

PROC DISCRIM

Q: How do I diplay/output discriminant function coefficients, linear or quadratic?
A: If METHOD=NORMAL and POOL=YES (the default, resulting in linear functions) or POOL=NO (resulting in quadratic functions), the coefficients are stored in OUTSTAT=. Also, when the functions are linear, there is a page of printed output labeled `Linear Discriminant Function` (like page 762, bottom). When METHOD=NPAR, no functions are used, and therefore they cannot be displayed or output. With METHOD=NPAR, an observation is classified by determining the group for which it has the maximum density (if using R=) or the largest number of the nearest neighbors (if K= is used). See pp. 682-683 of the *SAS/STAT User's Guide, Volume 1.*

Q: Why are there more discriminant functions than I expected?
A: People often expect that when they have 2 groups (classes) they should get 1 function. For classification purposes, this gets very cumbersome when there are >2 classes. Therefore, PROC DISCRIM always creates as many functions as classes. However, if you really want to get only 1 function for 2 classes, try PROC CANDISC. It will give one fewer functions than classes, but the functions cannot be easily used for classification. See "Classifying Observations — How to" in *SAS/STAT User's Guide, Volume 1.*

Q: How do I classify observations?
A: Given the discriminant functions, all you do is take the dot product of the observation with each of the functions. Whichever group's function gives the largest result is the group into which the observations are classified. But you do not have to do this yourself unless you don't want to use SAS. Example 5 on pages 760-761 in *SAS/STAT User's Guide, Volume 1*, (example 2, page 330 in the Version 5 book) shows how to use the TESTDATA= option to let the procedure do the work. The general method is also discussed on page 698 (pp. 322-323 in the Version 5 book).

Q: But why do I classify according to the LARGEST value? Aren't the dot product values the distances of the observations to the group means?
A: No. The dot product values are just scores. They are related to the distances, but are different, partly, by a factor of −2. As a result, maximizing these scores minimizes the distances.

Q: Which variables are most important (are best discriminators)?
A: Use PROC STEPDISC to help select the best discriminating variables. Once a set of variables is selected, PROC CANDISC (or PROC DISCRIM with the CANONICAL option) can tell you how important each variable is via the structure coefficients. The discriminant function coefficients from PROC DISCRIM are **NOT** the best way to determine variable importance because correlation among the variables can make their interpretation in that way misleading.

Q: Can I use categorical variables?
A: The normal theory method (METHOD=NORMAL, the default) assumes multivariate normality. Using categorical variables violates this assumption. Logistic modeling is a better, and simpler, approach. Use PROC LOGISTIC for a two-group situation or PROC CATMOD for more than two groups. If data sparsity causes problems with PROC CATMOD or PROC LOGISTIC, or if discriminant analysis is required, try the nonparametric methods available with METHOD=NPAR. Then use either K= for the *k*-nearest-neighbor method or R= for the kernel density method.

PROC STEPDISC

Q: Can PROC STEPDISC print the discriminant functions?
A: No, use PROC DISCRIM.

Q: Can PROC STEPDISC output the selected variables to discriminate?
A: No.

Q: Which variables are most important (are best discriminators)?
A: Use PROC STEPDISC to help select the best discriminating variables. Once a set of variables is selected, PROC CANDISC (or PROC DISCRIM with the CANONICAL option) can tell you how important each variable is via the structure coefficients. The discriminant function coefficients from PROC DISCRIM are NOT the best way to determine variable importance since correlation among the variables can make their interpretation in that way misleading.

SAS/STAT SOFTWARE—REGRESSION:

PROC REG

Q: The NOINT option in the MODEL statement causes a note that the R-square has been redefined. How is it defined with NOINT?
A: The R-square is still calculated as the regression sum of squares divided by the total sum of squares. However, when there is not an intercept in the model, we use uncorrected, rather than corrected, sums of squares. The best reference for exactly how the calculations are being performed is the section on multiple correlation in *Linear Models* (Searle 1971, pp. 95-98, John Wiley & Sons, Inc.). A good discussion of regression through the origin is also found in Section 2.4.5 (pages 33-38) of the *SAS System for Regression, Second Edition*. Note that PROC GLM calculates the R-square for the NOINT model in the same way. For a discussion of the properties of different estimators of R-square see Kvalseth's paper, "A Cautionary Note About R-square," in the *American Statistician* (November, 1985). It discusses eight different estimators of R-square, all of which are equivalent for a linear model with an intercept, in terms of both linear models without an intercept and nonlinear models. The most important things to note are that the R-square for a nointercept model is calculated differently from that for a linear model with an intercept and that the two should NOT be directly compared. For those wanting to make such a comparison, Myers suggests calculating an R-square for the NOINT model as:

1 - (error ss from noint model / corrected total ss).

For further discussion, see pp. 30-31 of Myers' book *Classical and Modern Regression with Applications* (1986, PWS Publishers).

Q: How do I get my parameter estimates/beta values/regression coefficients into a SAS data set?
A: The OUTEST= option in the PROC REG statement creates a data set that contains the parameter estimates and optionally the variance-covariance matrix of the parameter estimates (if you also specify the COVOUT option in the PROC REG statement). If you want the t-statistics and probabilities, see the REGCI sample library program. See pages 1389-1391 of the *SAS/STAT User's Guide*, for a discussion of what is available in this data set. Many enhancements to this data set are documented on page 526 of SAS Technical Report P-229 (for example, you can now get the R-square, adjusted R-square, and many other stats in this data set).

Q: How can I get the standardized parameter estimates into the OUTEST data set?
A: Use PROC STANDARD to standardize the data first (PROC STANDARD MEAN=0 STD=1;), then run PROC REG with the OUTEST= option specified in the PROC statement. When the data is standardized prior to going into PROC REG, the regular estimates are the same as those produced by the STB option when PROC REG is executed on the raw data.

Q: How can I get PROC REG to use the standard errors produced by the ACOV option in calculating my *t*-statistics?
A: If you specify single degree of freedom tests (using the TEST statement), PROC REG will compute the test using the regular standard errors and then will recompute the test using the standard errors from the ACOV matrix. It will be a chi-square, rather than a t, because the standard errors are asymptotic, rather than exact. See pp. 820-821 of Hal White's *Econometrica* paper for details (May, 1980).

Q: Can I apply an inequality constraint (for example, B1>0) to the parameters in PROC REG?
A: The RESTRICT statement in PROC REG allows only equality constraints. However, you can use the BOUNDS statement in PROC NLIN to apply inequality constraints. NLIN solves linear models quickly!

Q: What exactly is the SPEC option testing?
A: The SPEC option is testing a joint null hypothesis of the following:

the errors are homoskedastic

the errors are independent of the regressors

the model is correctly specified

Failure of any of these conditions can lead to a statistically significant test statistic. Essentially, the statistic is testing a joint hypothesis that the model's specification of the first and second moments of the dependent variable is correct. See the bottom of page 823 of Hal White's May 1980 *Econometrica* paper for more details (the SPEC option is Theorem 2).

Q: How does PROC REG/PROC GLM/and so on compute this or that statistic?
A: See Chapter 1, "Introduction to Regression Procedures," in the *SAS/STAT User's Guide* for many formulas (confidence limits, standard errors, and so on) that are used by **ALL** regression procedures.

SAS/STAT SOFTWARE—SURVIVAL PROCEDURES:

PROC LIFEREG

Q: How do I get the shape and scale parameters for a Weibull distribution from PROC LIFEREG?
A: PROC LIFEREG will print out a scale and an intercept parameter — but these are from an accelerated failure time model (see "Introduction" in the PROC LIFEREG documentation for details). The intercept is called mu, and the scale is called sigma on the PROC LIFEREG output. These parameters may be transformed to produce the Weibull parameters alpha and gamma, as shown in the "Distributions Allowed" section. Gamma=1/sigma and alpha=exp(-mu/sigma). Note that different authors use different forms of the probability density function, and so you might have to do a bit of algebra to make g(t) look like the pdf with which you are familiar.

PROC LIFETEST

1st available: Release 6.03 update (doc.: SAS Technical Report P-179)

Q: How can I output (insert any statistic here) to a data set?
A: If it can't be output in the OUTTEST= or OUTSURV= data sets, there is an experimental version of PROC LIFETEST available that can use the output delivery system (ODS) in Release 6.08 on VMS, OS/2, and Windows.

PROC PHREG

1st available: Release 6.07.01 (doc.: P-229; Release 6.04 by request. doc.: SAS Technical Report P-217)

Q: Is it available on the PC under Release 6.04?
A: Yes. The SAS site representative must call to request it, and it is free. It is documented in Technical Report P-217.

Q: How do I fit a multinomial logistic or discrete choice or McFadden's conditional logistic model?
A: See "Multinomial Logit, Discrete Choice Modeling" by Warren Kuhfeld. It is available as TS-273.

MISCELLANEOUS ENTRIES:

DISTANCE macro

1st available: Release 6.09 sample library (doc.: internal; see header)

Q: What does it require?
A: Just base SAS, unless the STD=AGK(p) or STD=L(p) options are used, in which case SAS/STAT is needed. If you want to standardize variables (STD= option) or replace missing values (MISSING=, OPTIONS=REPLACE, or OPTIONS=REPONLY), then the STDIZE macro is needed too. You must run on Release 6.06 or higher.

Q: What does it do?
A: The %DISTANCE macro computes various measures of distance, dissimilarity, or similarity between the observations (rows) of a SAS data set. These proximity measures are stored as a lower triangular matrix in an output data set that can then be used as input to the CLUSTER, MDS, or MODECLUS procedures. The input data set may contain numeric or character variables or both, depending on which proximity measure is used.

Q: How do I get it?

A: You can get these files via anonymous ftp:

```
get
/techsup/download/samples/stat_and_iml/distance.sas
and /techsup/download/samples/stat_and_iml/xmacro.sas
```

and (optionally):

```
/techsup/download/samples/stat_and_iml/stdize.sas
```

SIBBS:

```
stat distance.sas stat xmacro.sas stat stdize.sas
```

To use, see documentation in headers.

FUNCTIONS (BASE - PROBABILITY, RANDOM NUMBER, QUANTILE)

Q: Is there a function that will give the value of PI?
A: There is no PI() function, but you can get it's value with either of the following: arcos(-1) or gamma(1/2)**2 . However, the precision is not unlimited. For example, on HP700 it is correct to 14 decimal places.

Q: Is there a function that will give the factorial of a number?
A: The GAMMA function can be used to obtain the factorial of a number. For positive integers, GAMMA(X) is (X-1)!. Thus, to find 6!, in a DATA step you would use an assignment statement such as:

```
sixfact = gamma(7);
```

Q: Are there functions for the inverses of COSH, SINH, and TANH (the hyperbolic cosine, sine, and tangent)?
A: No, but they are very easy to obtain using existing SAS functions in the DATA step as follows:

```
inv(cosh): arcosh_x=log(x+sqrt(x**2-1)); inv(sinh):
arsinh_x=log(x+sqrt(x**2+1)); inv(tanh):
artanh_x=0.5*log((1+x)/(1-x));
```

(where X is the value from COSH, SINH, or TANH, respectively.) Inverses can be obtained similarly for COTH, SECH, and CSCH. See *SAS Communications* 4th quarter 1993.

Q: Is there a function that can be used to determine the number of potential combinations of n objects, selected r at a time (this problem is often referred to as combinatorics)?
A: Because of the number of potential combinations of n objects, selected r at a time is given by: n! / (r! * (n-r)!). We could use the GAMMA function as discussed above, but this will only work when n is relatively small (because the computations can overflow rather quickly). The exact limit before the GAMMA function overflows is machine dependent (it is, for example, greater than about 57 (56!) on an IBM® mainframe). However, another approach would be to use the EXP and LGAMMA functions. In general,

A! B! / C! D! =
EXP(LGAMMA(A+1)+LGAMMA(B+1)-LGAMMA(C+1)-LGAMMA(D+1)).

Thus, the number of ways of selecting two things at a time out of a total of five is given by:

5! / (2! 3!) = 10

and can be computed either as:

METHOD1 = GAMMA(6) / (GAMMA(3) * GAMMA(4));
or METHOD2 = EXP(LGAMMA(6)-LGAMMA(3)-LGAMMA(4)).

INTRACC MACRO

Q: What is required?
A: Base SAS and SAS/STAT.

Q: What does it do?
A: This macro calculates the six intraclass correlations discussed in Shrout and Fleiss "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, 1979, 86, 420-428. Additionally, it calculates two intraclass correlations using formulae from Winer, *Statistical Principles in Experimental Design*, (1962, McGraw-Hill). Additionally, it calculates the reliability of the mean of N ratings where N is a parameter of the macro, using the Spearman-Brown prophecy formula, to examine the effect that obtaining more raters would have on the reliability of a mean.

ITEM MACRO

1st available: Release 6.07.03 sample library (doc.: internal; see header) (there was a Version 5 supplemental procedure called PROC ITEM)

Q: What is required?
A: Base SAS, Release 6.04 or later. If running this under DOS, you will probably need expanded memory.

Q: What does it do?
A: The %ITEM macro computes descriptive statistics for analysis of data from a multiple-choice test. Each observation contains the answers from one subject to a set of questions (items). The data are compared to an answer key to determine which answers are correct. The score for each subject is computed as the number of correct answers. The output is very similar to that from the ITEM procedure in the SUGI Supplemental library, but several incorrect statistics have been fixed.

NEURAL NETWORK MACROS: %TNN_NLP, %TNN_RUN (SAS/OR) and %TNN_MODL (ETS)

1st available: %TNN_NLP & %TNN_RUN: Release 6.09 SAS/OR sample lib.; member NLPEX7 %TNN_MODL: 6.10 ETS sample lib., member MODEL7 (doc.: internal; see headers)

Q: What do they do?
A: These macros are intended as EXAMPLES of how artificial neural networks can be implemented with SAS software. These macros are **NOT** full-featured production-quality programs suitable for serious data analysis in a wide range of practical applications. The %TNN_NLP macro fits a standard feed-forward neural network with one hidden layer (a multilayer perceptron) as a nonlinear multivariate regression model by least squares using the SAS/OR procedure NLP for nonlinear programming. The %TNN_SIM macro simulates a network using parameter estimates (weights) previously computed by the %TNN_NLP macro. The predicted values and residuals are stored in the OUT= data set. Various statistics measuring goodness or badness of fit are stored in the OUTFIT= data set. The %TNN_MODL model does the same thing as the %TNN_NLP and %TNN_RUN macros but uses the SAS/ETS MODEL procedure instead of PROC NLP in SAS/OR.

Q: What do they require?
A: SAS/OR Release 6.08 or higher needs PROC NLP. %TNN_MODL does not require SAS/OR, but does require SAS/ETS (PROC MODEL).

Q: How do I get them?

A: You can get them via anonymous ftp:
`get /techsup/download/samples/stat_and_iml/nlpex7.sas`
or `/techsup/download/samples/stat_and_iml/model7.sas`
SIBBS:
stat `nlpex7.sas`
stat `model7.sas`

PROC RANK

Q: What percentile/quartile/decile is each observation in?
A: Use the RANK procedure (documented in the *SAS Procedures Guide, Version 6, Third Edition*) for this. Specify GROUPS=100 to get percentile ranks. GROUPS=10 for decile ranks, and GROUPS=4 for quartile ranks.

STDIZE macro

1st available: Release 6.09 sample library (doc.: internal; see header)

Q: What does it require?
A: Just base SAS unless the METHOD=AGK(p) or METHOD=L(p) options are used, in which case SAS/STAT is needed. You must use Release 6.06 or later.

Q: What does it do?
A: The %STDIZE macro standardizes one or more numeric variables in a SAS data set by subtracting a location measure and dividing by a scale measure. A variety of location and scale measures are provided, including estimates that are resistant to outliers and clustering (see the METHOD= argument). You can also multiply each standardized value by a constant and add a constant.

Q: How do I get it?

A: You can get these files via anonymous ftp:
`get /techsup/download/samples/stat_and_iml/xmacro.sas`
and `/techsup/download/samples/stat_and_iml/stdize.sas`
SIBBS:
stat `xmacro.sas`
stat `stdize.sas`

To use, see documentation in headers.

PROC UNIVARIATE

Q: Can I get just the box plot? Just the probability plot? Just the histogram?
A: No, the plots cannot be selected separately. If you have SAS/QC, PROC CAPABILITY can be used to create probability plots and histograms, and PROC SHEWHART will produce box plots. SAS/GRAPH can also be used to get box plots with PROC GPLOT and the INTERPOL=BOX option in a SYMBOL statement. See SAS Technical Report A-106 for code to produce probability plots using only base SAS. See TS-50 for probability plots using SAS/GRAPH.

Q: I'm trying to test my data for normality with the NORMAL option in PROC UNIVARIATE. How do I interpret the values that I get?
A: You can always look at the probability value (labeled either PROB<W or PROB>D). Small values indicate non-normality. It is up to you to determine what small is but typical choices are 0.05 and 0.10. Refer to the article on page 38 of the Second Quarter 1990 *SAS Communications* and page 627 of the *SAS Procedures Guide*. Be sure and look at plots given with the PLOT option too. The normality test will almost always be rejected with large sample sizes due to the power of the test, but it might still be close enough to normality for practical purposes. If the sample size is big but the plot looks normal enough, you might want to take that into consideration too.

Q: Can UNIVARIATE compute a weighted median or other quantile?
A: No. The WEIGHT statement does not affect the quantiles. If the weight values are not extremely large and are integers (or can be made integers by multiplying them by a constant that doesn't make them too large), then the FREQ statement can be used. A weighted median can be computed with FASTCLUS in Release 6.07.03 or later. The Cluster Median reported by FASTCLUS is a weighted median when the options LEAST=1 (documented in SAS Technical Report P-229) and MAXCLUSTERS=1 are used along with a WEIGHT statement. The statistic computed by FASTCLUS is a median in the sense that it minimizes the weighted sum of the p-th powers of the absolute deviations.

Q: How does PROC MEANS/PROC SUMMARY/PROC UNIVARIATE compute this or that statistic?
A: The answer is probably in Chapter 1 of the *SAS Procedures Guide* ("SAS Elementary Statistics Procedures"— see page 4, in particular). There is also a very useful table on pages 2-3 that shows which descriptive statistics are computed in which procedures.

Q: My standard deviation (or variance) computed by PROC MEANS/PROC SUMMARY/ PROC UNIVARIATE is much too large. What is the matter?
A: Check to see if a WEIGHT statement is being used. If so, you probably also need to use the VARDEF= option in the procedure statement (you probably want to use VARDEF=WGT but that is up to you).

Q: What percentile/quartile/decile is each observation in?
A: Use the RANK procedure (documented in the *SAS Procedures Guide*) for this. Specify GROUPS=100 to get percentile ranks, GROUPS=10 for decile ranks, and GROUPS=4 for quartile ranks.

Q: Why don't I get the same p-value for the Kolmogorov-Smirnov statistic when I use a table?
A: The critical values of the K-S statistic depend on N (as with most statistics). However, Stephens (reference is given at end of the PROC UNIVARIATE chapter) came up with a clever transformation of the K-S statistic that incorporates N so that you can compare the transformed statistic against a single set of critical values. That is what we do in SAS/INSIGHT and PROC UNIVARIATE.

STATISTICAL ITEMS THAT USERS FREQUENTLY REQUEST

Test or Item	PROC/MODULE/SAMPLE Program
AID (AUTOMATIC INTERACTIQN DETECTOR)	See CHAID below
ALPHA (CRONBACH'S)	CORR
ANOVA USING SUMMARY STATISTICS	DATA step, refer to article in *American Statistician*, May 1992, by Larson
ARCH (AUTOREGRESSIVE COND'L HETEROSCEDASTICITY)	AUTOREG (Release 6.08)
AREA ESTIMATION	TS-233
BALANCED INCOMPLETE BLOCK DESIGNS (BIBDs)	Beginning in Release 6.10 there are additional options in OPTEX for creating BIBD designs. For earlier releases, see SAS Technical Report P-188, Chapter 6, example 6 using PROC OPTEX. We can't guarantee that OPTEX will find BIBDs using this pre-Release 6.10 method, but it often does.
BARTLETT (TEST FOR VARIANCE HOMOGENEITY)	SAMPLE LIBRARY - BARTLETT. Or download `homovar.sas` from STAT area of SIBBS. Or download `techsup/download/samples/stat_and_iml/homovar.sas` using anonymous ftp.
BHAPKAR'S TEST	CATMOD using REPEATED statement to test marginal homogeneity as in example 7. Reference is Agresti (1990), *Categorical Data Analysis*, pp. 359, 499.
BONFERRONI T-TEST	GLM, ANOVA, MULTTEST
BOX AND WHISKER PLOTS	UNIVARIATE, SHEWHART, GPLOT (I=BOX... IN SYMBOL STMT) AND PROC INSIGHT (Release 6.08)
BOX-COX TRANSFORMATIONS	ADX menu system or ADXTRANS macro in SAS/QC *ONLY* (there is nothing else to do this).
CART	We don't do it currently, but we are considering a procedure to do it. See CHAID below.
CHAID (CHI-SQUARE AUTOMATIC INTERACTION DETECTOR)	A macro (%TREEDISC) to do this will be in the Release 6.10 SAS/IML sample library. Earlier beta testing has occurred. The macro will be sent to all customers on request. It is also available on anonmyous ftp and must be accompanied by by XMACRO.SAS.
CHI-SQUARE (GOODNESS OF FIT, 1-WAY TABLES)	SAMPLE LIBRARY (**6.07.01 AND LATER**) (MEMBER: PEARSON) *Observations* article (Fourth Quarter 1991, p. 17) *SAS Communications* articles (90Q1 p. 43 & 91Q3 p. 35) TS-176 U6062219, U6041468, U5186436
CHI-SQUARE (2-WAY TABLES)	FREQ
CHI-SQUARE(CORRECTED)	FREQ
CLUSTER ANALYSIS, CATEGORICAL DATA	...see DISTANCES
COCHRAN'S Q	FREQ (CMH OPTION). Done in same way as McNemar's test. See "Example", p. 129 of the *SAS/STAT User's Guide, Volume 1.*
CONDITIONAL LOGISTIC MODEL	For 1:1 ("one to one") matching, see example 5 in LOGISTIC. For M:N matching (for M>1, N>1), see example 3 in PHREG documentation. Or, see Multinomial Logit model.
CONFIDENCE INTERVAL (ON A MEAN)	DATA step, MEANS (Release 6.07.03)
CONJOINT ANALYSIS	Technical Report R-109: Conjoint Analysis Examples. Available from Book Sales.
CONTRASTS	GLM, CATMOD
CONTROL CHARTS	SHEWHART, MACONTROL, CUSUM
CORRELATIONS	...CORR (PEARSON, SPEARMAN, KENDALL)
CPK (PROCESS CAPABILITY INDICES)	CAPABILITY
CPM (CRITICAL PATH METHOD	CPM
D (SOMER'S)	FREQ
DECISION TREES	DTREE
DIFFERENTIAL EQUATIONS	IML (CALL ODE)
DISCRETE CHOICE MODEL	(See Multinomial Logit Model)

DISTANCES ..... %DISTANCE MACRO in DISTANCE member of the Release 6.09 SAS/STAT sample library.
DUNCAN MULTIPLE RANGE TEST ..... GLM, ANOVA
DUNNETT'S TEST ..... GLM, ANOVA
DURBIN-WATSON STATISTIC ..... REG, AUTOREG
DXY AND DYX(SOMER'S) ..... FREQ
EMPIRICAL DISTRIBUTION FUNCTION ..... EDF option in NPAR1WAY
EQUALITY OF MEANS ..... ANOVA, GLM, TTEST
FISHBONE DIAGRAMS ..... ISHIKAWA
FISHER'S EXACT TEST ..... FREQ
FISHER'S LEAST SIG. DIFFERENCE ..... ANOVA, GLM
FRIEDMAN'S TEST ..... FREQ (If only one response per treatment-block combination) or PROC RANK; BY BLOCK, THEN GLM
FULL-INFORMATION MAX. LIKELIHOOD ..... SYSLIN
GAM (Generalized Additive Models) ..... not available
GARCH (GENERALIZED AUTOREGRESSIVE COND'L HETEROSCEDASTICITY) ..... AUTOREG (Release 6.08)
GEE (Generalized Estimating Equations) ..... not available
GEOMETRIC MEANS ..... SAMPLE LIBRARY (MEMBER: GEOMEAN)
GINI'S MEAN DIFFERENCE ..... INSIGHT
HOEFFDING'S D ..... CORR
HOMOGENEITY OF VARIANCE ..... SAMPLE LIBRARY (MEMBER: BARTLETT)
HOTELLING'S T-SQUARE ..... *SAS System for Linear Models*, pp. 252-255.
INTEGRATION ..... IML (CALL QUAD IN Release 6.07.03) —- Also see area estimation above
INTRACLASS CORRELATION ..... PROC NESTED (% OF TOTAL COLUMN) or %intracc macro An example using a data set in *American Statistician*, Vol 47, No. 4, page 293, is available on request.
INVERSE MILL'S RATIO (See MILL'S RATIO)
ISHIKAWA DIAGRAMS ..... ISHIKAWA (PROC Release 6.07.03 & later) ISHIKAWA MENUS & MACROS
ITEM ANALYSIS ..... SUGI PROC ITEM (IBM Version 5 ONLY) SAMPLE LIBRARY (MEMBER: ITEM) IN Release 6.07.03 AND LATER
JONCKHEERE'S TEST ..... PROC CORR WITH KENDALL OPTION (See SUGI 89, pp. 1337-1339)
KAPPA (COHEN'S, WEIGHTED OR NOT) ..... CATMOD (As shown in TS-118)
KENDALL CORRELATION ..... CORR
KENDALL'S TAU A ..... FREQ
KURTOSIS ..... UNIVARIATE, MEANS
LD50 ..... PROC PROBIT W/ INVERSECL OPTION
LEVENE'S TEST (VARIANCES) ..... SAS/LAB (Release 6.08, 6.07 beta)
LIKELIHOOD RATIO CHI-SQUARE ..... FREQ
LINEAR PROGRAMMING (OPTIMIZATION) ..... LP
LJUNG-BOX Q STATISTIC ..... ARIMA
MANTEL-HAENSZEL CHI SQUARE ..... FREQ
McFadden's model ..... See Conditional Logistic Model.
MCNEMAR'S TEST ..... FREQ (CMH OPTION). See example on p. 129, *SAS/STAT User's Guide, Volume 1.*
MEDIAN ..... UNIVARIATE
MILL'S RATIO (OR INVERSE) ..... We do not compute it, but if you want it in the context of a probit model, there is a user-contributed program that will do it. See PROBIT section of statfaqs file. In addition, it is on our list of suggestions.
MOVING AVERAGE ..... PROC EXPAND (SAS/ETS) in Release 6.08 and higher supports this with the TRANSFORM=option in the CONVERT statement. Point users to 2nd Edition of *SAS/ETS User's Guide*. If they don't have SAS/ETS then Sample Library (member: MAVERAGE). Also, see pp. 223-228 of *SAS Language and Procedures, Usage 2.*
MLOGIT or MPROBIT procedures ..... These are not procedures written or supported by SAS Institute. If you want a Multinomial Logit or Discrete-Choice model or Mill's ratio, see the entries for these. See Version 5 SAS Usage Note 3639 concerning these proceduress.
MULTINOMIAL LOGIT MODEL ..... PHREG — Send TS-273
MULTIPLE COMPARISONS(REPEATED MEAS) TS-235
NEURAL NETWORKS ..... %TNN_NLP in NLPEX7 member of Release 6.09 SAS/OR sample library.
NORMALITY, TEST OF ..... see TEST OF NORMALITY below
NUMERICAL INTEGRATION ..... IML (CALL QUAD)
ODDS RATIOS ..... LOGISTIC (Release 6.07.03 and later)
OPTIMIZATION ..... LP, NETFLOW, NLP, TRANS, ASSIGN
OUTLIER DETECTION ..... See *SAS System for Statistical Graphics*, Chapter 9.
PAIRED COMPARISONS T-TEST ..... MEANS (See Example. 2 in TTEST DOC.)
PARETO CHARTS ..... PARETO
PARTIAL LEAST SQUARES ..... We currently have nothing for this. They will have to write an IML program.
PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS (PBIBDs) ..... See Balanced Incomplete Block Designs above. There is nothing that specifically creates these, but OPTEX might do it in some cases.
PATH ANALYSIS ..... CALIS
PEARSON CORRELATION ..... CORR
PERCENTILES ..... UNIVARIATE, PCTL(SUGI)
PERT ..... CPM
PHI COEFFICIENT ..... FREQ
POINT BISERIAL CORRELATIONS ..... CORR (See Release 6.06 SAS Usage Note 3956)
POLYCHORIC CORRELATION COEFFICIENT ..... FREQ PLCORR option in TABLES statement, Release 6.07.03 and up. See SAS Technical Report P-222, pg 223.
POLYSERIAL CORRELATIONS ..... NOT AVAILABLE
POWER ..... TECHNICAL NOTES TS-89,TS-90,TS-91 NEW SUGI PAPER, SUGI17 pp.1173-1177 (This will replace the above papers soon.)
PROBABILITY PLOTS ..... CAPABILITY
PROJECT MANAGEMENT ..... CPM, PROJMAN
QUEUEING ..... See SIMULATION below
RANDOM SAMPLING ..... DATA step — See the *SAS APPLICATIONS GUIDE* or the *Language and Procedures: Usage 2.*
REGRESSION ..... REG, GLM
RELIABILITY COEFFICIENT (SPSS) ..... See ALPHA (above)
REPEATED MEASURES ..... MULTIPLE COMPARISONS - TS-235
ROC (Receiver Operating Characteristic) CURVE ..... LOGISTIC - In Release 6.10, use the OUTROC= option. Before Release 6.10, plot sensitivity against 100-specificity values from output of CTABLE option after specifying PPROB=0 to 1 by .05 .
RSQUARE ..... REG, GLM, CALIS
SCHEFFE MULTIPLE COMPARISONS ..... GLM, ANOVA
SEEMINGLY UNRELATED REGRESSION ..... SYSLIN, MODEL
SIMULATION ..... NOT CURRENTLY AVAILABLE - please track to SAS/STAT or other
SKEWNESS ..... UNIVARIATE, MEANS
SOMER'S D ..... FREQ
SOMER'S DXY AND DYX ..... FREQ
SPEARMAN CORRELATION ..... CORR

STANDARDIZING DATA..... %STDIZE macro in STDIZE member of the Release 6.09 SAS/STAT sample library
STUART-MAXWELL TEST..... CATMOD using REPEATED statement to test marginal homogeneity as in example 7. Reference is Agresti (1990), *Categorical Data Analysis*, pp. 359, 499.
STUDENT-NEWMAN-KEULS TEST..... GLM, ANOVA
STUDENTS T-TEST..... UNIVARIATE, MEANS, SUMMARY
SUM OF SQUARES.....UNIVARIATE, MEANS, SUMMARY, GLM
SUMMARY STATISTICS-ANOVA..... See ANOVA USING SUMMARY STATISTICS
TRIMMED MEAN..... INSIGHT
T-TEST..... TTEST
T-TEST(BONFERRONI)..... GLM, ANOVA, MULTTEST
TAU-A(KENDALL'S)..... FREQ
TAU-C(STUART'S)..... FREQ
TEST OF HOMOGENEITY OF VARIANCE.....SAMPLE LIBRARY (MEMBER: BARTLETT)
TEST OF NORMALITY..... UNIVARIATE, CAPABILITY. We have no test for multivariate normality.
TETRACHORIC CORRELATIONS..... PLCORR option in FREQ (Release 6.07.03 or later — option documented in P-229). The tetrachoric correlation is a polychoric correlation for 2x2 tables.
THREE-STAGE LEAST SQUARES (3SLS)..... SYSLIN, MODEL
TREND TEST FOR 2 BY C TABLES..... PROC MULTTEST'S (Release 6.07.03) COCHRAN-ARMITAGE TEST; PROC LOGISTIC's score test; PROC FREQ with CMH and SCORES=rank option
TOLERANCE INTERVALS..... CAPABILITY
TUKEY'S RANGE TEST..... GLM, ANOVA
TWO-STAGE LEAST SQUARES (2SLS)..... SYSLIN, MODEL
WEIGHTED MEANS..... MEANS
WINSORIZED MEAN..... INSIGHT

EXTRA INFORMATION

ANONYMOUS FTP

SAS Institute's anonymous ftp service can be used to upload or download files without prior registration.

To use this facility, connect to `ftp.sas.com`. Once you are connected, enter the following responses as prompted:

```
Name (ftp.sas.com:<userid>): anonymous
Password: <your E-Mail address>
```

If a you get a message that `ftp.sas.com` is an unknown host, use the IP address instead; the IP address is 192.35.83.8.

SIBBS

The phone number to access SIBBS is: (919)677-8155. Communication software must be configured for no parity, 8 data bits, and 1 stop bit (8-N-1). Speed up to 14,400 baud.

To register, log in as follows:

```
First and LAST name:   first and last name
City, State:           city and state
PASSWORD:              whatever password you desire
```

You will be prompted to enter your specific information. Accounts will be available for use immediately.

You can track problems, review problems, resolve or add to them. You can request maintenance and submit SASware Ballots® and also have the capability of downloading zaps and SAS Usage Notes.

SAS, SAS/ETS, SAS/FSP, SAS/GRAPH, SAS/IML, SAS/INSIGHT, SAS/OR, SAS/QC, SAS/STAT, *SAS Communications*, the SASware Ballot, and *Observations* are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. IBM is a registered trademark or trademark of International Business Machines Corporation. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.