

SPSS Advanced Models™ 12.0



For more information about SPSS® software products, please visit our Web site at <http://www.spss.com> or contact

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412
Tel: (312) 651-3000
Fax: (312) 651-3668

SPSS is a registered trademark and the other product names are the trademarks of SPSS Inc. for its proprietary computer software. No material describing such software may be produced or distributed without the written permission of the owners of the trademark and license rights in the software and the copyrights in the published materials.

The SOFTWARE and documentation are provided with RESTRICTED RIGHTS. Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subdivision (c) (1) (ii) of The Rights in Technical Data and Computer Software clause at 52.227-7013. Contractor/manufacturer is SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

General notice: Other product names mentioned herein are used for identification purposes only and may be trademarks of their respective companies.

TableLook is a trademark of SPSS Inc.

Windows is a registered trademark of Microsoft Corporation.

DataDirect, DataDirect Connect, INTERSOLV, and SequeLink are registered trademarks of DataDirect Technologies.

Portions of this product were created using LEADTOOLS © 1991-2000, LEAD Technologies, Inc. ALL RIGHTS RESERVED.

LEAD, LEADTOOLS, and LEADVIEW are registered trademarks of LEAD Technologies, Inc.

Portions of this product were based on the work of the FreeType Team (<http://www.freetype.org>).

SPSS Advanced Models™ 12.0

Copyright © 2003 by SPSS Inc.

All rights reserved.

Printed in the United States of America.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Preface

SPSS 12.0 is a powerful software package for microcomputer data management and analysis. The Advanced Models option is an add-on enhancement that provides additional statistical analysis techniques. The procedures in Advanced Models must be used with the SPSS 12.0 Base and are completely integrated into that system.

The Advanced Models option includes procedures for:

- Multivariate and repeated measures general linear models
- Variance components analysis
- Linear mixed models
- Loglinear analysis, including model selection (hierarchical), general loglinear models, and logit loglinear models
- Ordinal regression analysis
- Survival analysis, including life tables, Kaplan-Meier survival analysis, and Cox regression

Installation

To install Advanced Models, follow the instructions for adding and removing features in the installation instructions supplied with the SPSS Base. (To start, double-click on the SPSS Setup icon.)

Compatibility

SPSS is designed to run on many computer systems. See the materials that came with your system for specific information on minimum and recommended requirements.

Serial Numbers

Your serial number is your identification number with SPSS Inc. You will need this serial number when you contact SPSS Inc. for information regarding support, payment, or an upgraded system. The serial number was provided with your Base system.

Customer Service

If you have any questions concerning your shipment or account, contact your local office, listed on the SPSS Web site at <http://www.spss.com/worldwide>. Please have your serial number ready for identification.

Training Seminars

SPSS Inc. provides both public and onsite training seminars. All seminars feature hands-on workshops. Seminars will be offered in major cities on a regular basis. For more information on these seminars, contact your local office, listed on the SPSS Web site at <http://www.spss.com/worldwide>.

Technical Support

The services of SPSS Technical Support are available to registered customers. Customers may contact Technical Support for assistance in using SPSS products or for installation help for one of the supported hardware environments. To reach Technical Support, see the SPSS Web site at <http://www.spss.com>, or contact your local office, listed on the SPSS Web site at <http://www.spss.com/worldwide>. Be prepared to identify yourself, your organization, and the serial number of your system.

Additional Publications

Additional copies of SPSS product manuals may be purchased directly from SPSS Inc. Visit our Web site at <http://www.spss.com/estore>, or contact your local SPSS office, listed on the SPSS Web site at <http://www.spss.com/worldwide>. For telephone orders in the United States and Canada, call SPSS Inc. at 800-543-2185. For telephone orders outside of North America, contact your local office, listed on the SPSS Web site.

The *SPSS Statistical Procedures Companion*, by Marija Norušis, is being prepared for publication by Prentice Hall. It contains overviews of the procedures in the SPSS Base, plus Logistic Regression, General Linear Models, and Linear Mixed Models. Further information will be available on the SPSS Web site at <http://www.spss.com> (click Store, select your country, and click Books).

Tell Us Your Thoughts

Your comments are important. Please send us a letter and let us know about your experiences with SPSS products. We especially like to hear about new and interesting applications using the SPSS system. Please send e-mail to suggest@spss.com or write to SPSS Inc., Attn.: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

About This Manual

This manual documents the graphical user interface for the procedures included in the Advanced Models module. Illustrations of dialog boxes are taken from SPSS for Windows. Dialog boxes in other operating systems are similar. The Advanced Models command syntax is included in the *SPSS 12.0 Command Syntax Reference*, available on the product CD-ROM.

Contacting SPSS

If you would like to be on our mailing list, contact one of our offices, listed on our Web site at <http://www.spss.com/worldwide>. We will send you a copy of our newsletter and let you know about SPSS Inc. activities in your area.

Contents

1	<i>GLM Multivariate Analysis</i>	1
	GLM Multivariate Data Considerations	2
	Obtaining GLM Multivariate Tables	3
2	<i>GLM Repeated Measures</i>	17
	GLM Repeated Measures Data Considerations	18
	GLM Repeated Measures Define Factors	20
	Defining Factors for GLM Repeated Measures	21
	Obtaining GLM Repeated Measures Tables	22
3	<i>Variance Components Analysis</i>	35
	Variance Components Data Considerations	36
	Obtaining Variance Components Tables	36
4	<i>Linear Mixed Models</i>	43
	Linear Mixed Models Data Considerations	44
	Linear Mixed Models Select Subjects/Repeated Variables	44
	Selecting Subjects/Repeated Variables for Linear Mixed Models	46

	Obtaining a Linear Mixed Models Analysis	47
5	<i>Model Selection Loglinear Analysis</i>	59
	Model Selection Loglinear Analysis Data Considerations	59
	Obtaining a Model Selection Loglinear Analysis	60
6	<i>General Loglinear Analysis</i>	65
	General Loglinear Analysis Data Considerations	66
	Obtaining a General Loglinear Analysis	66
7	<i>Logit Loglinear Analysis</i>	73
	Logit Loglinear Analysis Data Considerations	74
	Obtaining a Logit Loglinear Analysis	74
8	<i>Ordinal Regression</i>	81
	Ordinal Regression Data Considerations	82
	Obtaining an Ordinal Regression	82
9	<i>Life Tables</i>	91
	Life Tables Data Considerations	92

Creating Life Tables	93
10 Kaplan-Meier Survival Analysis	97
Kaplan-Meier Data Considerations	97
Obtaining a Kaplan-Meier Survival Analysis	98
11 Cox Regression Analysis	105
Cox Regression Data Considerations	105
Obtaining a Cox Regression Analysis	106
12 Computing Time-Dependent Covariates	113
Computing a Time-Dependent Covariate	114
 Appendices	
A Categorical Variable Coding Schemes	117
Deviation	117
Simple	118
Helmert	119
Difference	120
Polynomial	120

Repeated	121
Special	122
Indicator	123

B Covariance Structures **125**

Index **131**

GLM Multivariate Analysis

The GLM Multivariate procedure provides regression analysis and analysis of variance for multiple dependent variables by one or more factor variables or covariates. The factor variables divide the population into groups. Using this general linear model procedure, you can test null hypotheses about the effects of factor variables on the means of various groupings of a joint distribution of dependent variables. You can investigate interactions between factors as well as the effects of individual factors. In addition, the effects of covariates and covariate interactions with factors can be included. For regression analysis, the independent (predictor) variables are specified as covariates.

Both balanced and unbalanced models can be tested. A design is balanced if each cell in the model contains the same number of cases. In a multivariate model, the sums of squares due to the effects in the model and error sums of squares are in matrix form rather than the scalar form found in univariate analysis. These matrices are called SSCP (sums-of-squares and cross-products) matrices. If more than one dependent variable is specified, the multivariate analysis of variance using Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root criterion with approximate F statistic are provided as well as the univariate analysis of variance for each dependent variable. In addition to testing hypotheses, GLM Multivariate produces estimates of parameters.

Commonly used *a priori* contrasts are available to perform hypothesis testing. Additionally, after an overall F test has shown significance, you can use post hoc tests to evaluate differences among specific means. Estimated marginal means give estimates of predicted mean values for the cells in the model, and profile plots (interaction plots) of these means allow you to visualize some of the relationships easily. The post hoc multiple comparison tests are performed for each dependent variable separately.

Residuals, predicted values, Cook's distance, and leverage values can be saved as new variables in your data file for checking assumptions. Also available are a residual SSCP matrix, which is a square matrix of sums of squares and cross-products of residuals, a residual covariance matrix, which is the residual SSCP matrix divided by the degrees of freedom of the residuals, and the residual correlation matrix, which is the standardized form of the residual covariance matrix.

WLS Weight allows you to specify a variable used to give observations different weights for a weighted least-squares (WLS) analysis, perhaps to compensate for different precision of measurement.

Example. A manufacturer of plastics measures three properties of plastic film: tear resistance, gloss, and opacity. Two rates of extrusion and two different amounts of additive are tried, and the three properties are measured under each combination of extrusion rate and additive amount. The manufacturer finds that the extrusion rate and the amount of additive individually produce significant results but that the interaction of the two factors is not significant.

Methods. Type I, Type II, Type III, and Type IV sums of squares can be used to evaluate different hypotheses. Type III is the default.

Statistics. Post hoc range tests and multiple comparisons: least significant difference, Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiple F , Ryan-Einot-Gabriel-Welsch multiple range, Student-Newman-Keuls, Tukey's honestly significant difference, Tukey's- b , Duncan, Hochberg's GT2, Gabriel, Waller Duncan t test, Dunnett (one-sided and two-sided), Tamhane's T2, Dunnett's T3, Games-Howell, and Dunnett's C . Descriptive statistics: observed means, standard deviations, and counts for all of the dependent variables in all cells; the Levene test for homogeneity of variance; Box's M test of the homogeneity of the covariance matrices of the dependent variables; and Bartlett's test of sphericity.

Plots. Spread-versus-level, residual, and profile (interaction).

GLM Multivariate Data Considerations

Data. The dependent variables should be quantitative. Factors are categorical and can have numeric values or string values of up to eight characters. Covariates are quantitative variables that are related to the dependent variable.

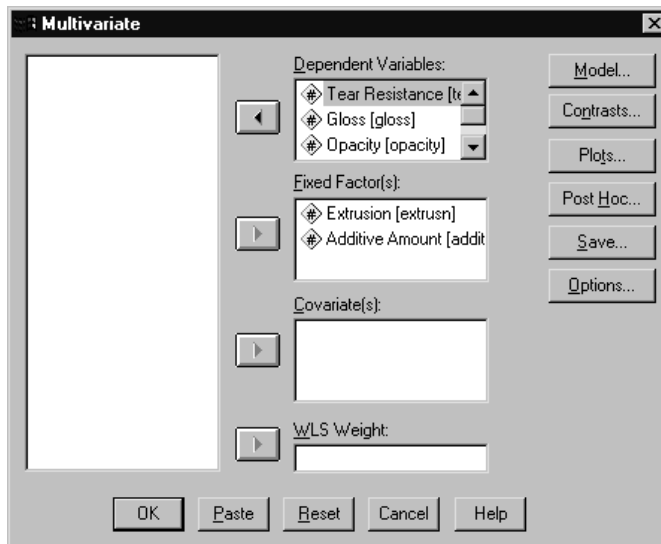
Assumptions. For dependent variables, the data are a random sample of vectors from a multivariate normal population; in the population, the variance-covariance matrices for all cells are the same. Analysis of variance is robust to departures from normality, although the data should be symmetric. To check assumptions, you can use homogeneity of variances tests (including Box's M) and spread-versus-level plots. You can also examine residuals and residual plots.

Related procedures. Use the Explore procedure to examine the data before doing an analysis of variance. For a single dependent variable, use GLM Univariate. If you measured the same dependent variables on several occasions for each subject, use GLM Repeated Measures.

Obtaining GLM Multivariate Tables

- ▶ From the menus choose:
Analyze
 General Linear Model
 Multivariate...

Figure 1-1
Multivariate dialog box

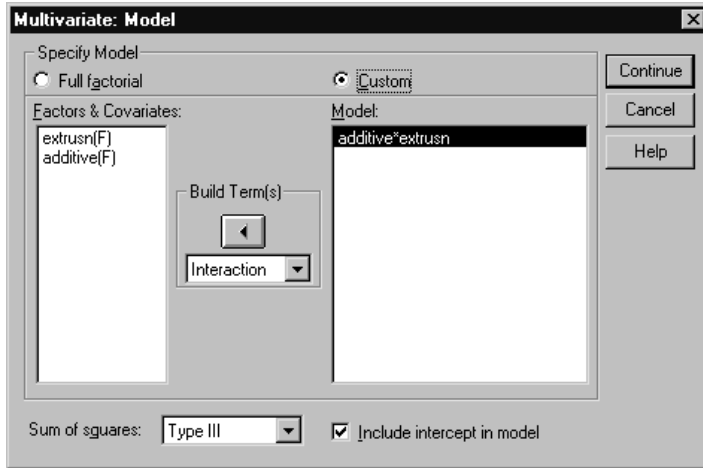


- ▶ Select at least two dependent variables.

Optionally, you can specify Fixed Factor(s), Covariate(s), and WLS Weight.

GLM Multivariate Model

Figure 1-2
Multivariate Model dialog box



Specify Model. A full factorial model contains all factor main effects, all covariate main effects, and all factor-by-factor interactions. It does not contain covariate interactions. Select Custom to specify only a subset of interactions or to specify factor-by-covariate interactions. You must indicate all of the terms to be included in the model.

Factors and Covariates. The factors and covariates are listed with (F) for fixed factor and (C) for covariate.

Model. The model depends on the nature of your data. After selecting Custom, you can select the main effects and interactions that are of interest in your analysis.

Sum of squares. The method of calculating the sums of squares. For balanced or unbalanced models with no missing cells, the Type III sum-of-squares method is the most commonly used type.

Include intercept in model. The intercept is usually included in the model. If you can assume that the data pass through the origin, you can exclude the intercept.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Sums of Squares

For the model, you can choose a type of sum of squares. Type III is the most commonly used and is the default.

Type I. This method is also known as the hierarchical decomposition of the sum-of-squares method. Each term is adjusted only for the term that precedes it in the model. The Type I sum-of-squares method is commonly used for:

- A balanced ANOVA model in which any main effects are specified before any first-order interaction effects, any first-order interaction effects are specified before any second-order interaction effects, and so on.
- A polynomial regression model in which any lower-order terms are specified before any higher-order terms.
- A purely nested model in which the first-specified effect is nested within the second-specified effect, the second-specified effect is nested within the third, and so on. (This form of nesting can be specified only by using syntax.)

Type II. This method calculates the sums of squares of an effect in the model adjusted for all other “appropriate” effects. An appropriate effect is one that corresponds to all effects that do not **contain** the effect being examined. The Type II sum-of-squares method is commonly used for:

- A balanced ANOVA model.
- Any model that has main factor effects only.

- Any regression model.
- A purely nested design. (This form of nesting can be specified by using syntax.)

Type III. This method, the default, calculates the sums of squares of an effect in the design as the sums of squares adjusted for any other effects that do not contain it and orthogonal to any effects (if any) that contain it. The Type III sums of squares have one major advantage in that they are invariant with respect to the cell frequencies as long as the general form of estimability remains constant. Therefore, this type is often considered useful for an unbalanced model with no missing cells. In a factorial design with no missing cells, this method is equivalent to the Yates' weighted-squares-of-means technique. The Type III sum-of-squares method is commonly used for:

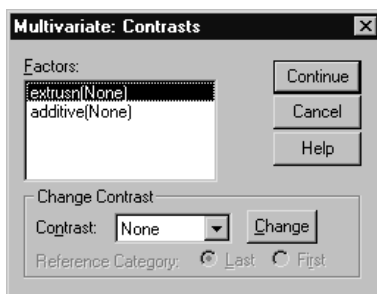
- Any models listed in Type I and Type II.
- Any balanced or unbalanced model with no empty cells.

Type IV. This method is designed for a situation in which there are missing cells. For any effect F in the design, if F is not contained in any other effect, then Type IV = Type III = Type II. When F is contained in other effects, Type IV distributes the contrasts being made among the parameters in F to all higher-level effects equitably. The Type IV sum-of-squares method is commonly used for:

- Any models listed in Type I and Type II.
- Any balanced model or unbalanced model with empty cells.

GLM Multivariate Contrasts

Figure 1-3
Multivariate Contrasts dialog box



Contrasts are used to test whether the levels of an effect are significantly different from one another. You can specify a contrast for each factor in the model. Contrasts represent linear combinations of the parameters.

Hypothesis testing is based on the null hypothesis $\mathbf{LBM} = \mathbf{0}$, where \mathbf{L} is the contrast coefficients matrix, \mathbf{M} is the identity matrix, which has dimension equal to the number of dependent variables, and \mathbf{B} is the parameter vector. When a contrast is specified, SPSS creates an \mathbf{L} matrix such that the columns corresponding to the factor match the contrast. The remaining columns are adjusted so that the \mathbf{L} matrix is estimable.

In addition to the univariate test using F statistics and the Bonferroni-type simultaneous confidence intervals based on Student's t distribution for the contrast differences across all dependent variables, the multivariate tests using Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root criteria are provided.

Available contrasts are deviation, simple, difference, Helmert, repeated, and polynomial. For deviation contrasts and simple contrasts, you can choose whether the reference category is the last or first category.

Contrast Types

Deviation. Compares the mean of each level (except a reference category) to the mean of all of the levels (grand mean). The levels of the factor can be in any order.

Simple. Compares the mean of each level to the mean of a specified level. This type of contrast is useful when there is a control group. You can choose the first or last category as the reference.

Difference. Compares the mean of each level (except the first) to the mean of previous levels. (Sometimes called reverse Helmert contrasts.)

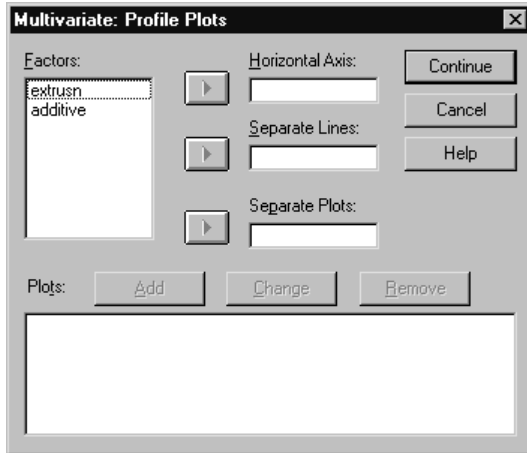
Helmert. Compares the mean of each level of the factor (except the last) to the mean of subsequent levels.

Repeated. Compares the mean of each level (except the last) to the mean of the subsequent level.

Polynomial. Compares the linear effect, quadratic effect, cubic effect, and so on. The first degree of freedom contains the linear effect across all categories; the second degree of freedom, the quadratic effect; and so on. These contrasts are often used to estimate polynomial trends.

GLM Multivariate Profile Plots

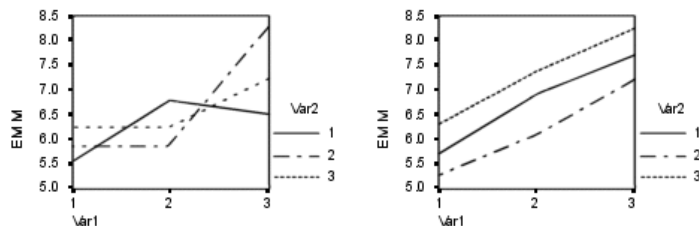
Figure 1-4
Multivariate Profile Plots dialog box



Profile plots (interaction plots) are useful for comparing marginal means in your model. Profile plots are created for each dependent variable. A profile plot is a line plot in which each point indicates the estimated marginal mean of a dependent variable (adjusted for covariates) at one level of a factor. The levels of a second factor can be used to make separate lines. Each level in a third factor can be used to create a separate plot. All fixed factors are available for plots.

A profile plot of one factor shows whether the estimated marginal means are increasing or decreasing across levels. For two or more factors, parallel lines indicate that there is no interaction between factors, which means that you can investigate the levels of only one factor. Lines that cross each other indicate an interaction.

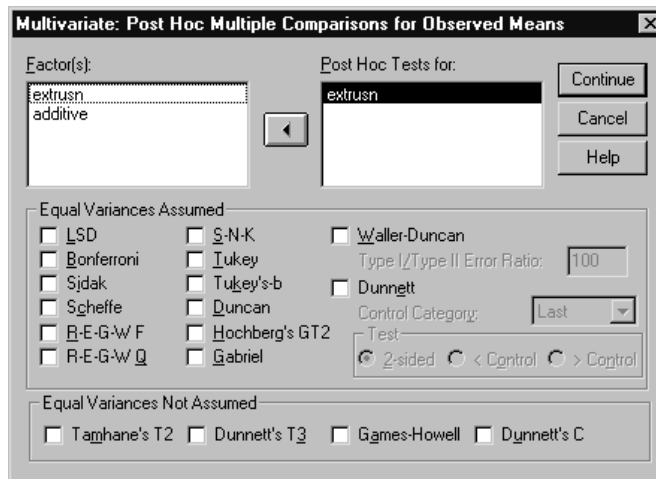
Figure 1-5
Nonparallel plot (left) and parallel plot (right)



After a plot is specified by selecting factors for the horizontal axis, and optionally, factors for separate lines and separate plots, the plot must be listed in the Plots list.

GLM Multivariate Post Hoc Multiple Comparisons for Observed Means

Figure 1-6
Post Hoc Multiple Comparisons for Observed Means dialog box



Post hoc multiple comparison tests. Once you have determined that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. Comparisons are made on unadjusted values. These tests are used for between-subjects factors only. The post hoc multiple comparison tests are performed for each dependent variable separately.

The Bonferroni and Tukey's honestly significant difference tests are commonly used multiple comparison tests. The **Bonferroni test**, based on Student's t statistic, adjusts the observed significance level for the fact that multiple comparisons are made. **Sidak's t test** also adjusts the significance level and provides tighter bounds than the Bonferroni test. **Tukey's honestly significant difference test** uses the Studentized range statistic to make all pairwise comparisons between groups and sets the experimentwise error rate to the error rate for the collection for all pairwise comparisons. When testing a large number of pairs of means, Tukey's honestly significant difference test is more powerful than the Bonferroni test. For a small number of pairs, Bonferroni is more powerful.

Hochberg's GT2 is similar to Tukey's honestly significant difference test, but the Studentized maximum modulus is used. Usually, Tukey's test is more powerful.

Gabriel's pairwise comparisons test also uses the Studentized maximum modulus and is generally more powerful than Hochberg's GT2 when the cell sizes are unequal. Gabriel's test may become liberal when the cell sizes vary greatly.

Dunnett's pairwise multiple comparison t test compares a set of treatments against a single control mean. The last category is the default control category. Alternatively, you can choose the first category. You can also choose a two-sided or one-sided test. To test that the mean at any level (except the control category) of the factor is not equal to that of the control category, use a two-sided test. To test whether the mean at any level of the factor is smaller than that of the control category, select < Control. Likewise, to test whether the mean at any level of the factor is larger than that of the control category, select > Control.

Ryan, Einot, Gabriel, and Welsch (R-E-G-W) developed two multiple step-down range tests. Multiple step-down procedures first test whether all means are equal. If all means are not equal, subsets of means are tested for equality. **R-E-G-W F** is based on an F test and **R-E-G-W Q** is based on the Studentized range. These tests are more powerful than Duncan's multiple range test and Student-Newman-Keuls (which are also multiple step-down procedures), but they are not recommended for unequal cell sizes.

When the variances are unequal, use **Tamhane's T2** (conservative pairwise comparisons test based on a t test), **Dunnett's T3** (pairwise comparison test based on the Studentized maximum modulus), **Games-Howell pairwise comparison test** (sometimes liberal), or **Dunnett's C** (pairwise comparison test based on the Studentized range).

Duncan's multiple range test, Student-Newman-Keuls (**S-N-K**), and **Tukey's-b** are range tests that rank group means and compute a range value. These tests are not used as frequently as the tests previously discussed.

The **Waller-Duncan t test** uses a Bayesian approach. This range test uses the harmonic mean of the sample size when the sample sizes are unequal.

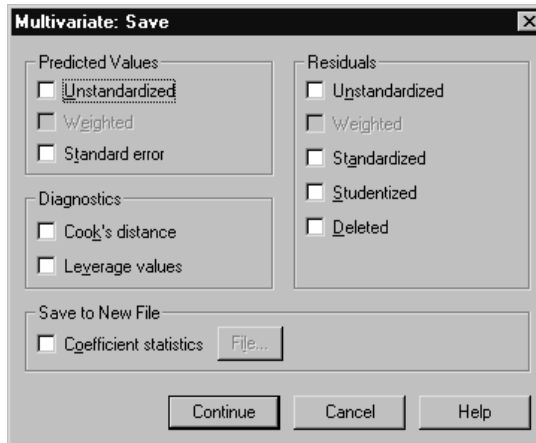
The significance level of the **Scheffé** test is designed to allow all possible linear combinations of group means to be tested, not just pairwise comparisons available in this feature. The result is that the Scheffé test is often more conservative than other tests, which means that a larger difference between means is required for significance.

The least significant difference (**LSD**) pairwise multiple comparison test is equivalent to multiple individual t tests between all pairs of groups. The disadvantage of this test is that no attempt is made to adjust the observed significance level for multiple comparisons.

Tests displayed. Pairwise comparisons are provided for LSD, Sidak, Bonferroni, Games and Howell, Tamhane's T2 and T3, Dunnett's C , and Dunnett's T3. Homogeneous subsets for range tests are provided for S-N-K, Tukey's- b , Duncan, R-E-G-W F , R-E-G-W Q , and Waller. Tukey's honestly significant difference test, Hochberg's GT2, Gabriel's test, and Scheffé's test are both multiple comparison tests and range tests.

GLM Multivariate Save

Figure 1-7
Multivariate Save dialog box



You can save values predicted by the model, residuals, and related measures as new variables in the Data Editor. Many of these variables can be used for examining assumptions about the data. To save the values for use in another SPSS session, you must save the current data file.

Predicted Values. The values that the model predicts for each case. Unstandardized predicted values and the standard errors of the predicted values are available. If a WLS variable was chosen, weighted unstandardized predicted values are available.

Diagnostics. Measures to identify cases with unusual combinations of values for the independent variables and cases that may have a large impact on the model. Available are Cook's distance and uncentered leverage values.

- **Cook's Distance.** A measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients. A large Cook's D indicates that excluding a case from computation of the regression statistics, changes the coefficients substantially.
- **Leverage Value.** The relative influence of each observation on the model's fit.

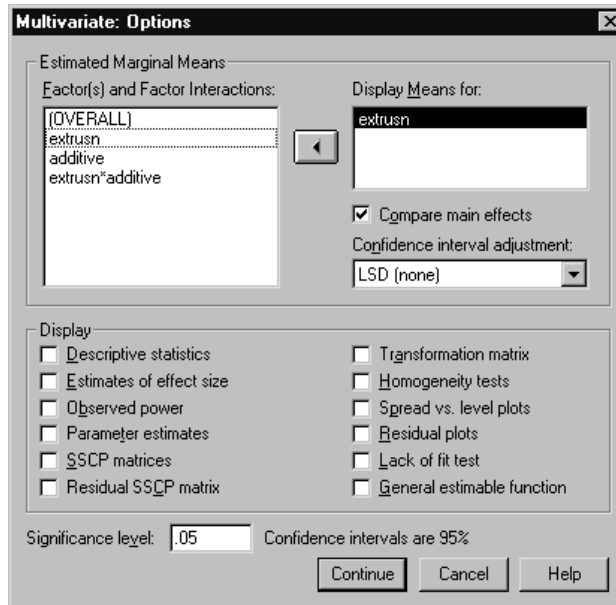
Residuals. An unstandardized residual is the actual value of the dependent variable minus the value predicted by the model. Standardized, Studentized, and deleted residuals are also available. If a WLS variable was chosen, weighted unstandardized residuals are available.

- **Unstandardized Residuals.** The difference between an observed value and the value predicted by the model.
- **Weighted Residuals.** Weighted unstandardized residuals. Available only if a WLS variable was previously selected.
- **Standardized.** The residual divided by an estimate of its standard error. Standardized residuals are also known as Pearson residuals.
- **Studentized Residual.** The residual divided by an estimate of its standard deviation that varies from case to case, depending on the distance of each case's values on the independent variables from the means of the independent variables.
- **Deleted Residual.** The residual for a case when that case is excluded from the calculation of the regression coefficients. It is the difference between the value of the dependent variable and the adjusted predicted value.

Save to New File. Writes an SPSS data file containing a variance-covariance matrix of the parameter estimates in the model. Also, for each dependent variable, there will be a row of parameter estimates, a row of significance values for the t statistics corresponding to the parameter estimates, and a row of residual degrees of freedom. For a multivariate model, there are similar rows for each dependent variable. You can use this data in other SPSS procedures.

GLM Multivariate Options

Figure 1-8
Multivariate Options dialog box



Optional statistics are available from this dialog box. Statistics are calculated using a fixed-effects model.

Estimated Marginal Means. Select the factors and interactions for which you want estimates of the population marginal means in the cells. These means are adjusted for the covariates, if any. Interactions are available only if you have specified a custom model.

- **Compare main effects.** Provides uncorrected pairwise comparisons among estimated marginal means for any main effect in the model, for both between- and within-subjects factors. This item is available only if main effects are selected under the Display Means For list.
- **Confidence interval adjustment.** Select least significant difference (LSD), Bonferroni, or Sidak adjustment to the confidence intervals and significance. This item is available only if Compare main effects is selected.

Display. Select Descriptive statistics to produce observed means, standard deviations, and counts for all of the dependent variables in all cells. Estimates of effect size gives a partial eta-squared value for each effect and each parameter estimate. The eta-squared statistic describes the proportion of total variability attributable to a factor. Select Observed power to obtain the power of the test when the alternative hypothesis is set based on the observed value. Select Parameter estimates to produce the parameter estimates, standard errors, *t* tests, confidence intervals, and the observed power for each test. You can display the hypothesis and error SSCP matrices and the Residual SSCP matrix plus Bartlett's test of sphericity of the residual covariance matrix.

Homogeneity tests produces the Levene test of the homogeneity of variance for each dependent variable across all level combinations of the between-subjects factors, for between-subjects factors only. Also, homogeneity tests include Box's *M* test of the homogeneity of the covariance matrices of the dependent variables across all level combinations of the between-subjects factors. The spread-versus-level and residual plots options are useful for checking assumptions about the data. This item is disabled if there are no factors. Select Residual plots to produce an observed-by-predicted-by-standardized residuals plot for each dependent variable. These plots are useful for investigating the assumption of equal variance. Select Lack of fit test to check if the relationship between the dependent variable and the independent variables can be adequately described by the model. General estimable function allows you to construct custom hypothesis tests based on the general estimable function. Rows in any contrast coefficient matrix are linear combinations of the general estimable function.

Significance level. You might want to adjust the significance level used in post hoc tests and the confidence level used for constructing confidence intervals. The specified value is also used to calculate the observed power for the test. When you specify a significance level, the associated level of the confidence intervals is displayed in the dialog box.

GLM Command Additional Features

These features may apply to univariate, multivariate, or repeated measures analysis. The SPSS command language also allows you to:

- Specify nested effects in the design (using the DESIGN subcommand).

- Specify tests of effects versus a linear combination of effects or a value (using the TEST subcommand).
- Specify multiple contrasts (using the CONTRAST subcommand).
- Include user-missing values (using the MISSING subcommand).
- Specify EPS criteria (using the CRITERIA subcommand).
- Construct a custom **L** matrix, **M** matrix, or **K** matrix (using the LMATRIX, MMATRIX, or KMATRIX subcommands).
- For deviation or simple contrasts, specify an intermediate reference category (using the CONTRAST subcommand).
- Specify metrics for polynomial contrasts (using the CONTRAST subcommand).
- Specify error terms for post hoc comparisons (using the POSTHOC subcommand).
- Compute estimated marginal means for any factor or factor interaction among the factors in the factor list (using the EMMEANS subcommand).
- Specify names for temporary variables (using the SAVE subcommand).
- Construct a correlation matrix data file (using the OUTFILE subcommand).
- Construct a matrix data file that contains statistics from the between-subjects ANOVA table (using the OUTFILE subcommand).
- Save the design matrix to a new data file (using the OUTFILE subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

GLM Repeated Measures

The GLM Repeated Measures procedure provides analysis of variance when the same measurement is made several times on each subject or case. If between-subjects factors are specified, they divide the population into groups. Using this general linear model procedure, you can test null hypotheses about the effects of both the between-subjects factors and the within-subjects factors. You can investigate interactions between factors as well as the effects of individual factors. In addition, the effects of constant covariates and covariate interactions with the between-subjects factors can be included.

In a doubly multivariate repeated measures design, the dependent variables represent measurements of more than one variable for the different levels of the within-subjects factors. For example, you could have measured both pulse and respiration at three different times on each subject.

The GLM Repeated Measures procedure provides both univariate and multivariate analyses for the repeated measures data. Both balanced and unbalanced models can be tested. A design is balanced if each cell in the model contains the same number of cases. In a multivariate model, the sums of squares due to the effects in the model and error sums of squares are in matrix form rather than the scalar form found in univariate analysis. These matrices are called SSCP (sums-of-squares and cross-products) matrices. In addition to testing hypotheses, GLM Repeated Measures produces estimates of parameters.

Commonly used *a priori* contrasts are available to perform hypothesis testing on between-subjects factors. Additionally, after an overall F test has shown significance, you can use post hoc tests to evaluate differences among specific means. Estimated marginal means give estimates of predicted mean values for the cells in the model, and profile plots (interaction plots) of these means allow you to visualize some of the relationships easily.

Residuals, predicted values, Cook's distance, and leverage values can be saved as new variables in your data file for checking assumptions. Also available are a residual SSCP matrix, which is a square matrix of sums of squares and cross-products of residuals, a residual covariance matrix, which is the residual SSCP matrix divided by the degrees of freedom of the residuals, and the residual correlation matrix, which is the standardized form of the residual covariance matrix.

WLS Weight allows you to specify a variable used to give observations different weights for a weighted least-squares (WLS) analysis, perhaps to compensate for different precision of measurement.

Example. Twelve students are assigned to a high- or low-anxiety group based on their scores on an anxiety-rating test. The anxiety rating is called a between-subjects factor because it divides the subjects into groups. The students are each given four trials on a learning task, and the number of errors for each trial is recorded. The errors for each trial are recorded in separate variables, and a within-subjects factor (trial) is defined with four levels for the four trials. The trial effect is found to be significant, while the trial-by-anxiety interaction is not significant.

Methods. Type I, Type II, Type III, and Type IV sums of squares can be used to evaluate different hypotheses. Type III is the default.

Statistics. Post hoc range tests and multiple comparisons (for between-subjects factors): least significant difference, Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiple F , Ryan-Einot-Gabriel-Welsch multiple range, Student-Newman-Keuls, Tukey's honestly significant difference, Tukey's- b , Duncan, Hochberg's GT2, Gabriel, Waller Duncan t test, Dunnett (one-sided and two-sided), Tamhane's T2, Dunnett's T3, Games-Howell, and Dunnett's C . Descriptive statistics: observed means, standard deviations, and counts for all of the dependent variables in all cells; the Levene test for homogeneity of variance; Box's M ; and Mauchly's test of sphericity.

Plots. Spread-versus-level, residual, and profile (interaction).

GLM Repeated Measures Data Considerations

Data. The dependent variables should be quantitative. Between-subjects factors divide the sample into discrete subgroups, such as male and female. These factors are categorical and can have numeric values or string values of up to eight characters. Within-subjects factors are defined in the Repeated Measures Define Factor(s)

dialog box. Covariates are quantitative variables that are related to the dependent variable. For a repeated measures analysis, these should remain constant at each level of a within-subjects variable.

The data file should contain a set of variables for each group of measurements on the subjects. The set has one variable for each repetition of the measurement within the group. A within-subjects factor is defined for the group with the number of levels equal to the number of repetitions. For example, measurements of weight could be taken on different days. If measurements of the same property were taken on five days, the within-subjects factor could be specified as *day* with five levels.

For multiple within-subjects factors, the number of measurements for each subject is equal to the product of the number of levels of each factor. For example, if measurements were taken at three different times each day for four days, the total number of measurements is 12 for each subject. The within-subjects factors could be specified as *day(4)* and *time(3)*.

Assumptions. A repeated measures analysis can be approached in two ways, univariate and multivariate.

The univariate approach (also known as the split-plot or mixed-model approach) considers the dependent variables as responses to the levels of within-subjects factors. The measurements on a subject should be a sample from a multivariate normal distribution, and the variance-covariance matrices are the same across the cells formed by the between-subjects effects. Certain assumptions are made on the variance-covariance matrix of the dependent variables. The validity of the F statistic used in the univariate approach can be assured if the variance-covariance matrix is circular in form (Huynh and Mandeville, 1979).

To test this assumption, Mauchly's test of sphericity can be used, which performs a test of sphericity on the variance-covariance matrix of an orthonormalized transformed dependent variable. Mauchly's test is automatically displayed for a repeated measures analysis. For small sample sizes, this test is not very powerful. For large sample sizes, the test may be significant even when the impact of the departure on the results is small. If the significance of the test is large, the hypothesis of sphericity can be assumed. However, if the significance is small and the sphericity assumption appears to be violated, an adjustment to the numerator and denominator degrees of freedom can be made in order to validate the univariate F statistic. Three estimates of this adjustment, which is called **epsilon**, are available in the GLM Repeated Measures procedure. Both the numerator and denominator degrees of freedom must be multiplied by epsilon, and the significance of the F ratio must be evaluated with the new degrees of freedom.

The multivariate approach considers the measurements on a subject to be a sample from a multivariate normal distribution, and the variance-covariance matrices are the same across the cells formed by the between-subjects effects. To test whether the variance-covariance matrices across the cells are the same, Box's M test can be used.

Related procedures. Use the Explore procedure to examine the data before doing an analysis of variance. If there are *not* repeated measurements on each subject, use GLM Univariate or GLM Multivariate. If there are only two measurements for each subject (for example, pre-test and post-test), and there are no between-subjects factors, you can use the Paired-Samples T Test procedure.

GLM Repeated Measures Define Factors

GLM Repeated Measures analyzes groups of related dependent variables that represent different measurements of the same attribute. This dialog box lets you define one or more within-subjects factors for use in GLM Repeated Measures. Note that the order in which you specify within-subjects factors is important. Each factor constitutes a level within the previous factor.

To use Repeated Measures, you must set up your data correctly. You must define within-subjects factors in this dialog box. Notice that these factors are not existing variables in your data but rather factors that you define here.

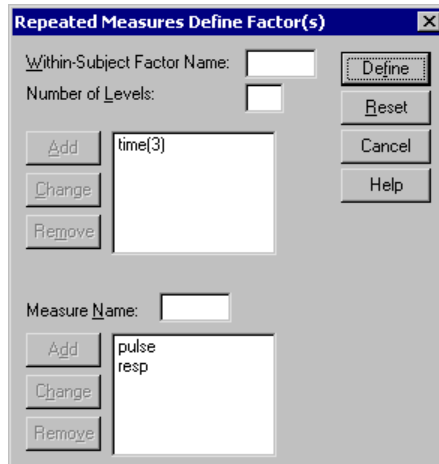
Example. In a weight-loss study, suppose the weights of several people are measured each week for five weeks. In the data file, each person is a subject or case. The weights for the weeks are recorded in the variables *weight1*, *weight2*, and so on. The gender of each person is recorded in another variable. The weights, measured for each subject repeatedly, can be grouped by defining a within-subjects factor. The factor could be called *week*, defined to have five levels. In the main dialog box, the variables *weight1*, ..., *weight5* are used to assign the five levels of *week*. The variable in the data file that groups males and females (*gender*) can be specified as a between-subjects factor to study the differences between males and females.

Measures. If subjects were tested on more than one measure at each time, define the measures. For example, the pulse and respiration rate could be measured on each subject every day for a week. These measures do not exist as variables in the data file but are defined here. A model with more than one measure is sometimes called a doubly multivariate repeated measures model.

Defining Factors for GLM Repeated Measures

- ▶ From the menus choose:
 - Analyze
 - General Linear Model
 - Repeated Measures...

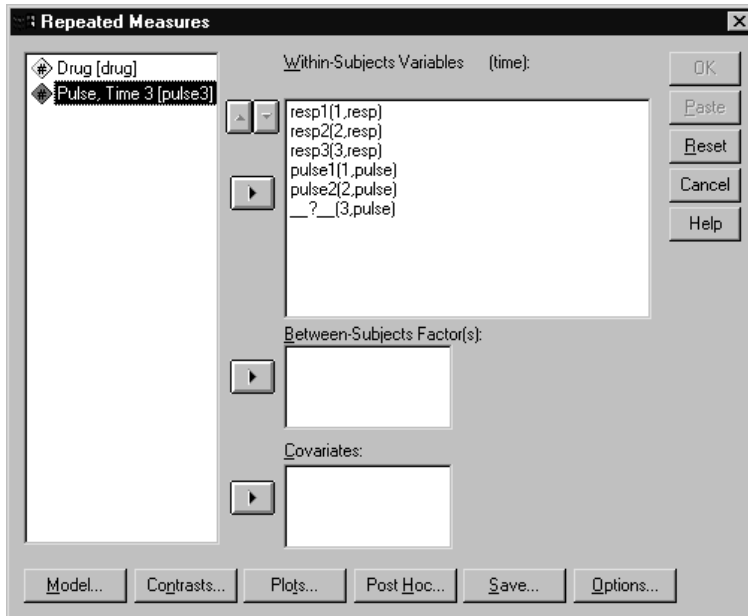
Figure 2-1
Repeated Measures Define Factor(s) dialog box



- ▶ Type a within-subject factor name and its number of levels.
- ▶ Click Add.
- ▶ Repeat these steps for each within-subjects factor.
 - To define measure factors for a doubly multivariate repeated measures design:
- ▶ Type the measure name.
- ▶ Click Add.
 - After defining all of your factors and measures:
- ▶ Click Define.

Obtaining GLM Repeated Measures Tables

Figure 2-2
Repeated Measures dialog box



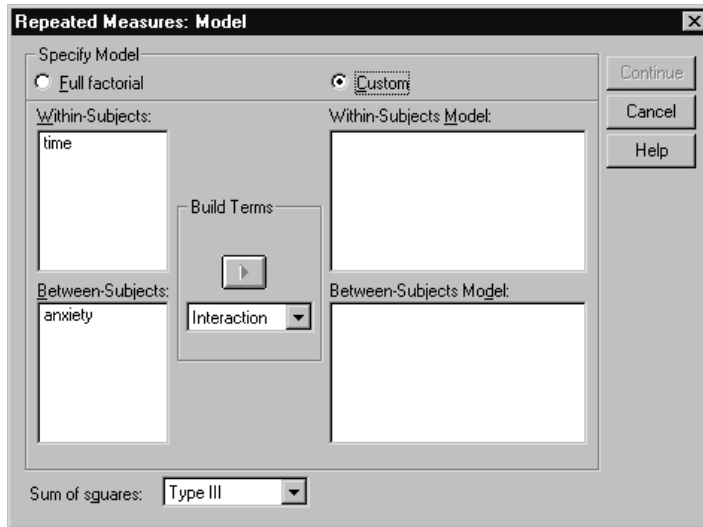
- ▶ Select a dependent variable that corresponds to each combination of within-subjects factors (and optionally, measures) on the list.

To change positions of the variables, use the up and down arrows.

To make changes to the within-subjects factors, you can reopen the Repeated Measures Define Factor(s) dialog box without closing the main dialog box. Optionally, you can specify between-subjects factor(s) and covariates.

GLM Repeated Measures Model

Figure 2-3
Repeated Measures Model dialog box



Specify Model. A full factorial model contains all factor main effects, all covariate main effects, and all factor-by-factor interactions. It does not contain covariate interactions. Select Custom to specify only a subset of interactions or to specify factor-by-covariate interactions. You must indicate all of the terms to be included in the model.

Between-Subjects. The covariates are listed with (C) for covariate.

Model. The model depends on the nature of your data. After selecting Custom, you can select the within-subjects effects and interactions and the between-subjects effects and interactions that are of interest in your analysis.

Sum of squares. The method of calculating the sums of squares for the between-subjects model. For balanced or unbalanced between-subjects models with no missing cells, the Type III sum-of-squares method is the most commonly used.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Sums of Squares

For the model, you can choose a type of sum of squares. Type III is the most commonly used and is the default.

Type I. This method is also known as the hierarchical decomposition of the sum-of-squares method. Each term is adjusted for only the term that precedes it in the model. The Type I sum-of-squares method is commonly used for:

- A balanced ANOVA model in which any main effects are specified before any first-order interaction effects, any first-order interaction effects are specified before any second-order interaction effects, and so on.
- A polynomial regression model in which any lower-order terms are specified before any higher-order terms.
- A purely nested model in which the first-specified effect is nested within the second-specified effect, the second-specified effect is nested within the third, and so on. (This form of nesting can be specified only by using syntax.)

Type II. This method calculates the sums of squares of an effect in the model adjusted for all other “appropriate” effects. An appropriate effect is one that corresponds to all effects that do not **contain** the effect being examined. The Type II sum-of-squares method is commonly used for:

- A balanced ANOVA model.
- Any model that has main factor effects only.

- Any regression model.
- A purely nested design. (This form of nesting can be specified by using syntax.)

Type III. This method, the default, calculates the sums of squares of an effect in the design as the sums of squares adjusted for any other effects that do not contain it and orthogonal to any effects (if any) that contain it. The Type III sums of squares have one major advantage in that they are invariant with respect to the cell frequencies as long as the general form of estimability remains constant. Therefore, this type is often considered useful for an unbalanced model with no missing cells. In a factorial design with no missing cells, this method is equivalent to the Yates' weighted-squares-of-means technique. The Type III sum-of-squares method is commonly used for:

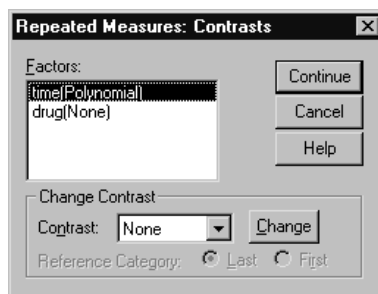
- Any models listed in Type I and Type II.
- Any balanced or unbalanced model with no empty cells.

Type IV. This method is designed for a situation in which there are missing cells. For any effect F in the design, if F is not contained in any other effect, then Type IV = Type III = Type II. When F is contained in other effects, Type IV distributes the contrasts being made among the parameters in F to all higher-level effects equitably. The Type IV sum-of-squares method is commonly used for:

- Any models listed in Type I and Type II.
- Any balanced model or unbalanced model with empty cells.

GLM Repeated Measures Contrasts

Figure 2-4
Repeated Measures Contrasts dialog box



Contrasts are used to test for differences among the levels of a between-subjects factor. You can specify a contrast for each between-subjects factor in the model. Contrasts represent linear combinations of the parameters.

Hypothesis testing is based on the null hypothesis $\mathbf{LBM}=0$, where \mathbf{L} is the contrast coefficients matrix, \mathbf{B} is the parameter vector, and \mathbf{M} is the average matrix that corresponds to the average transformation for the dependent variable. You can display this transformation matrix by selecting Transformation matrix in the Repeated Measures Options dialog box. For example, if there are four dependent variables, a within-subjects factor of four levels, and polynomial contrasts (the default) are used for within-subjects factors, the \mathbf{M} matrix will be $(0.5\ 0.5\ 0.5\ 0.5)'$. When a contrast is specified, SPSS creates an \mathbf{L} matrix such that the columns corresponding to the between-subjects factor match the contrast. The remaining columns are adjusted so that the \mathbf{L} matrix is estimable.

Available contrasts are deviation, simple, difference, Helmert, repeated, and polynomial. For deviation contrasts and simple contrasts, you can choose whether the reference category is the last or first category.

Contrast Types

Deviation. Compares the mean of each level (except a reference category) to the mean of all of the levels (grand mean). The levels of the factor can be in any order.

Simple. Compares the mean of each level to the mean of a specified level. This type of contrast is useful when there is a control group. You can choose the first or last category as the reference.

Difference. Compares the mean of each level (except the first) to the mean of previous levels. (Sometimes called reverse Helmert contrasts.)

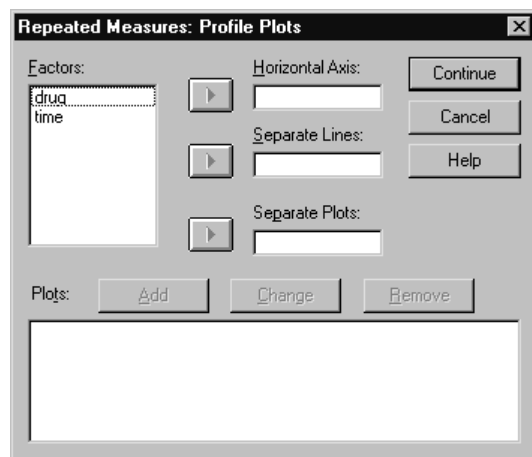
Helmert. Compares the mean of each level of the factor (except the last) to the mean of subsequent levels.

Repeated. Compares the mean of each level (except the last) to the mean of the subsequent level.

Polynomial. Compares the linear effect, quadratic effect, cubic effect, and so on. The first degree of freedom contains the linear effect across all categories; the second degree of freedom, the quadratic effect; and so on. These contrasts are often used to estimate polynomial trends.

GLM Repeated Measures Profile Plots

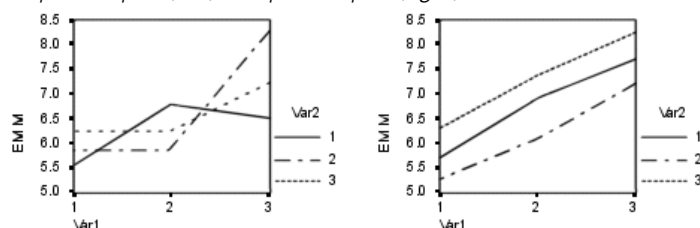
Figure 2-5
Repeated Measures Profile Plots dialog box



Profile plots (interaction plots) are useful for comparing marginal means in your model. A profile plot is a line plot in which each point indicates the estimated marginal mean of a dependent variable (adjusted for any covariates) at one level of a factor. The levels of a second factor can be used to make separate lines. Each level in a third factor can be used to create a separate plot. Both between-subjects factors and within-subjects factors can be used in profile plots.

A profile plot of one factor shows whether the estimated marginal means are increasing or decreasing across levels. For two or more factors, parallel lines indicate that there is no interaction between factors, which means that you can investigate the levels of only one factor. Nonparallel lines indicate an interaction.

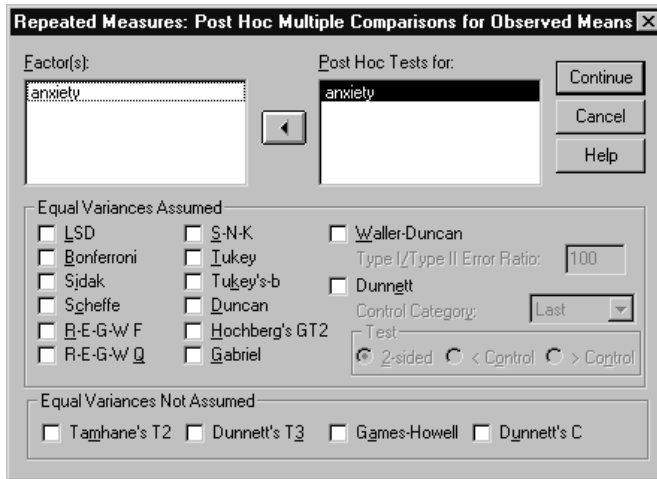
Figure 2-6
Nonparallel plot (left) and parallel plot (right)



After a plot is specified by selecting factors for the horizontal axis and, optionally, factors for separate lines and separate plots, it must be added to the Plots list.

GLM Repeated Measures Post Hoc Multiple Comparisons for Observed Means

Figure 2-7
Post Hoc Multiple Comparisons for Observed Means dialog box



Post hoc multiple comparison tests. Once you have determined that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. These tests are used for fixed *between-subjects* factors only. They are not available if there are no between-subjects factors. The post hoc multiple comparison tests are performed for the average across the levels of the within-subjects factors.

The Bonferroni and Tukey's honestly significant difference tests are commonly used multiple comparison tests. The **Bonferroni test**, based on Student's *t* statistic, adjusts the observed significance level for the fact that multiple comparisons are made. **Sidak's *t* test** also adjusts the significance level and provides tighter bounds than the Bonferroni test. **Tukey's honestly significant difference test** uses the Studentized range statistic to make all pairwise comparisons between groups and sets the experimentwise error rate to the error rate for the collection for all pairwise comparisons. When testing a large number of pairs of means, Tukey's honestly

significant difference test is more powerful than the Bonferroni test. For a small number of pairs, Bonferroni is more powerful.

Hochberg's GT2 is similar to Tukey's honestly significant difference test, but the Studentized maximum modulus is used. Usually, Tukey's test is more powerful.

Gabriel's pairwise comparisons test also uses the Studentized maximum modulus and is generally more powerful than Hochberg's GT2 when the cell sizes are unequal. Gabriel's test may become liberal when the cell sizes vary greatly.

Dunnett's pairwise multiple comparison t test compares a set of treatments against a single control mean. The last category is the default control category. Alternatively, you can choose the first category. You can also choose a two-sided or one-sided test. To test that the mean at any level (except the control category) of the factor is not equal to that of the control category, use a two-sided test. To test whether the mean at any level of the factor is smaller than that of the control category, select < Control. Likewise, to test whether the mean at any level of the factor is larger than that of the control category, select > Control.

Ryan, Einot, Gabriel, and Welsch (R-E-G-W) developed two multiple step-down range tests. Multiple step-down procedures first test whether all means are equal. If all means are not equal, subsets of means are tested for equality. **R-E-G-W F** is based on an F test and **R-E-G-W Q** is based on the Studentized range. These tests are more powerful than Duncan's multiple range test and Student-Newman-Keuls (which are also multiple step-down procedures), but they are not recommended for unequal cell sizes.

When the variances are unequal, use **Tamhane's T2** (conservative pairwise comparisons test based on a t test), **Dunnett's T3** (pairwise comparison test based on the Studentized maximum modulus), **Games-Howell pairwise comparison test** (sometimes liberal), or **Dunnett's C** (pairwise comparison test based on the Studentized range).

Duncan's multiple range test, Student-Newman-Keuls (**S-N-K**), and **Tukey's-b** are range tests that rank group means and compute a range value. These tests are not used as frequently as the tests previously discussed.

The **Waller-Duncan t test** uses a Bayesian approach. This range test uses the harmonic mean of the sample size when the sample sizes are unequal.

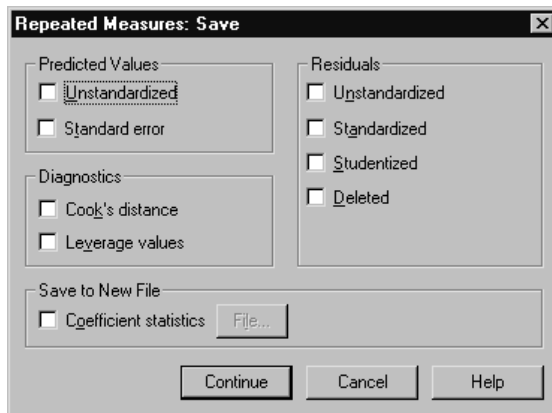
The significance level of the **Scheffé** test is designed to allow all possible linear combinations of group means to be tested, not just pairwise comparisons available in this feature. The result is that the Scheffé test is often more conservative than other tests, which means that a larger difference between means is required for significance.

The least significant difference (**LSD**) pairwise multiple comparison test is equivalent to multiple individual t tests between all pairs of groups. The disadvantage of this test is that no attempt is made to adjust the observed significance level for multiple comparisons.

Tests displayed. Pairwise comparisons are provided for LSD, Sidak, Bonferroni, Games and Howell, Tamhane's T2 and T3, Dunnett's C , and Dunnett's T3. Homogeneous subsets for range tests are provided for S-N-K, Tukey's- b , Duncan, R-E-G-W F , R-E-G-W Q , and Waller. Tukey's honestly significant difference test, Hochberg's GT2, Gabriel's test, and Scheffé's test are both multiple comparison tests and range tests.

GLM Repeated Measures Save

Figure 2-8
Repeated Measures Save dialog box



You can save values predicted by the model, residuals, and related measures as new variables in the Data Editor. Many of these variables can be used for examining assumptions about the data. To save the values for use in another SPSS session, you must save the current data file.

Predicted Values. The values that the model predicts for each case. Unstandardized predicted values and the standard errors of the predicted values are available.

Diagnostics. Measures to identify cases with unusual combinations of values for the independent variables and cases that may have a large impact on the model. Available are Cook's distance and uncentered leverage values.

- **Cook's Distance.** A measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients. A large Cook's D indicates that excluding a case from computation of the regression statistics, changes the coefficients substantially.
- **Leverage Value.** The relative influence of each observation on the model's fit.

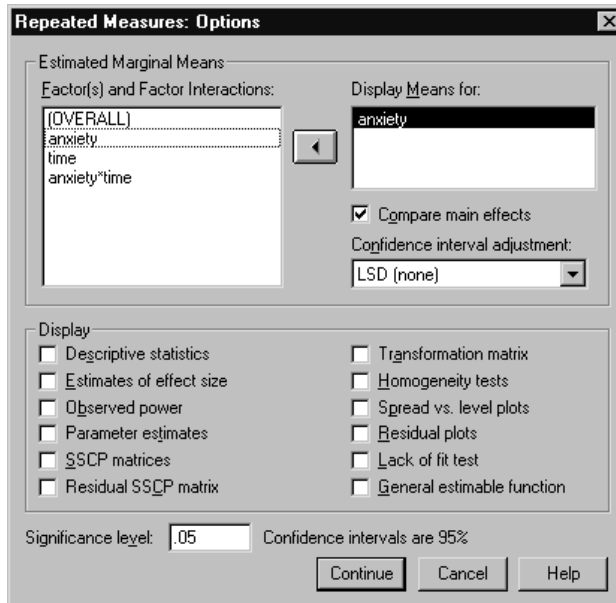
Residuals. An unstandardized residual is the actual value of the dependent variable minus the value predicted by the model. Standardized, Studentized, and deleted residuals are also available.

- **Unstandardized Residuals.** The difference between an observed value and the value predicted by the model.
- **Standardized.** The residual divided by an estimate of its standard error. Standardized residuals are also known as Pearson residuals.
- **Studentized Residual.** The residual divided by an estimate of its standard deviation that varies from case to case, depending on the distance of each case's values on the independent variables from the means of the independent variables.
- **Deleted Residual.** The residual for a case when that case is excluded from the calculation of the regression coefficients. It is the difference between the value of the dependent variable and the adjusted predicted value.

Save to New File. Writes an SPSS data file containing a variance-covariance matrix of the parameter estimates in the model. Also, for each dependent variable, there will be a row of parameter estimates, a row of significance values for the *t* statistics corresponding to the parameter estimates, and a row of residual degrees of freedom. For a multivariate model, there are similar rows for each dependent variable. You can use this matrix file in other procedures that read an SPSS matrix file.

GLM Repeated Measures Options

Figure 2-9
Repeated Measures Options dialog box



Optional statistics are available from this dialog box. Statistics are calculated using a fixed-effects model.

Estimated Marginal Means. Select the factors and interactions for which you want estimates of the population marginal means in the cells. These means are adjusted for the covariates, if any. Both within-subjects and between-subjects factors can be selected.

- **Compare main effects.** Provides uncorrected pairwise comparisons among estimated marginal means for any main effect in the model, for both between- and within-subjects factors. This item is available only if main effects are selected under the Display Means For list.
- **Confidence interval adjustment.** Select least significant difference (LSD), Bonferroni, or Sidak adjustment to the confidence intervals and significance. This item is available only if Compare main effects is selected.

Display. Select Descriptive statistics to produce observed means, standard deviations, and counts for all of the dependent variables in all cells. Estimates of effect size gives a partial eta-squared value for each effect and each parameter estimate. The eta-squared statistic describes the proportion of total variability attributable to a factor. Select Observed power to obtain the power of the test when the alternative hypothesis is set based on the observed value. Select Parameter estimates to produce the parameter estimates, standard errors, *t* tests, confidence intervals, and the observed power for each test. You can display the hypothesis and error SSCP matrices and the Residual SSCP matrix plus Bartlett's test of sphericity of the residual covariance matrix.

Homogeneity tests produces the Levene test of the homogeneity of variance for each dependent variable across all level combinations of the between-subjects factors, for between-subjects factors only. Also, homogeneity tests include Box's *M* test of the homogeneity of the covariance matrices of the dependent variables across all level combinations of the between-subjects factors. The spread-versus-level and residual plots options are useful for checking assumptions about the data. This item is disabled if there are no factors. Select Residual plots to produce an observed-by-predicted-by-standardized residuals plot for each dependent variable. These plots are useful for investigating the assumption of equal variance. Select Lack of fit test to check if the relationship between the dependent variable and the independent variables can be adequately described by the model. General estimable function allows you to construct custom hypothesis tests based on the general estimable function. Rows in any contrast coefficient matrix are linear combinations of the general estimable function.

Significance level. You might want to adjust the significance level used in post hoc tests and the confidence level used for constructing confidence intervals. The specified value is also used to calculate the observed power for the test. When you specify a significance level, the associated level of the confidence intervals is displayed in the dialog box.

GLM Command Additional Features

These features may apply to univariate, multivariate, or repeated measures analysis. The SPSS command language also allows you to:

- Specify nested effects in the design (using the DESIGN subcommand).

- Specify tests of effects versus a linear combination of effects or a value (using the TEST subcommand).
- Specify multiple contrasts (using the CONTRAST subcommand).
- Include user-missing values (using the MISSING subcommand).
- Specify EPS criteria (using the CRITERIA subcommand).
- Construct a custom **L** matrix, **M** matrix, or **K** matrix (using the LMATRIX, MMATRIX, and KMATRIX subcommands).
- For deviation or simple contrasts, specify an intermediate reference category (using the CONTRAST subcommand).
- Specify metrics for polynomial contrasts (using the CONTRAST subcommand).
- Specify error terms for post hoc comparisons (using the POSTHOC subcommand).
- Compute estimated marginal means for any factor or factor interaction among the factors in the factor list (using the EMMEANS subcommand).
- Specify names for temporary variables (using the SAVE subcommand).
- Construct a correlation matrix data file (using the OUTFILE subcommand).
- Construct a matrix data file that contains statistics from the between-subjects ANOVA table (using the OUTFILE subcommand).
- Save the design matrix to a new data file (using the OUTFILE subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

Variance Components Analysis

The Variance Components procedure, for mixed-effects models, estimates the contribution of each random effect to the variance of the dependent variable. This procedure is particularly interesting for analysis of mixed models such as split plot, univariate repeated measures, and random block designs. By calculating variance components, you can determine where to focus attention in order to reduce the variance.

Four different methods are available for estimating the variance components: minimum norm quadratic unbiased estimator (MINQUE), analysis of variance (ANOVA), maximum likelihood (ML), and restricted maximum likelihood (REML). Various specifications are available for the different methods.

Default output for all methods includes variance component estimates. If the ML method or the REML method is used, an asymptotic covariance matrix table is also displayed. Other available output includes an ANOVA table and expected mean squares for the ANOVA method, and an iteration history for the ML and REML methods. The Variance Components procedure is fully compatible with the GLM Univariate procedure.

WLS Weight allows you to specify a variable used to give observations different weights for a weighted analysis, perhaps to compensate for different precision of measurement.

Example. At an agriculture school, weight gains for pigs in six different litters are measured after one month. The litter variable is a random factor with six levels. (The six litters studied are a random sample from a large population of pig litters.) The investigator finds out that the variance in weight gain is attributable to the difference in litters much more than to the difference in pigs within a litter.

Variance Components Data Considerations

Data. The dependent variable is quantitative. Factors are categorical. They can have numeric values or string values of up to eight characters. At least one of the factors must be random. That is, the levels of the factor must be a random sample of possible levels. Covariates are quantitative variables that are related to the dependent variable.

Assumptions. All methods assume that model parameters of a random effect have zero means and finite constant variances and are mutually uncorrelated. Model parameters from different random effects are also uncorrelated.

The residual term also has a zero mean and finite constant variance. It is uncorrelated with model parameters of any random effect. Residual terms from different observations are assumed to be uncorrelated.

Based on these assumptions, observations from the same level of a random factor are correlated. This fact distinguishes a variance component model from a general linear model.

ANOVA and MINQUE do not require normality assumptions. They are both robust to moderate departures from the normality assumption.

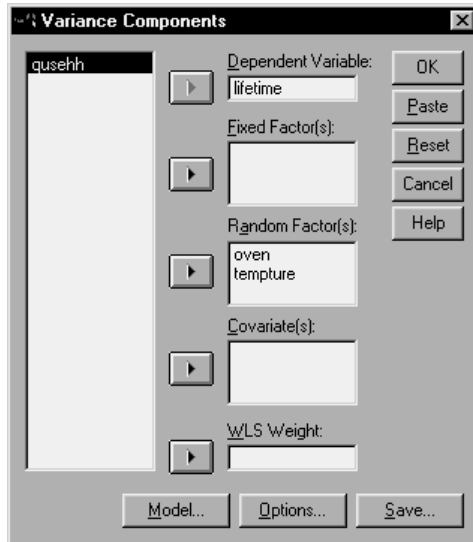
ML and REML require the model parameter and the residual term to be normally distributed.

Related procedures. Use the Explore procedure to examine the data before doing variance components analysis. For hypothesis testing, use GLM Univariate, GLM Multivariate, and GLM Repeated Measures.

Obtaining Variance Components Tables

- ▶ From the menus choose:
 - Analyze
 - General Linear Model
 - Variance Components...

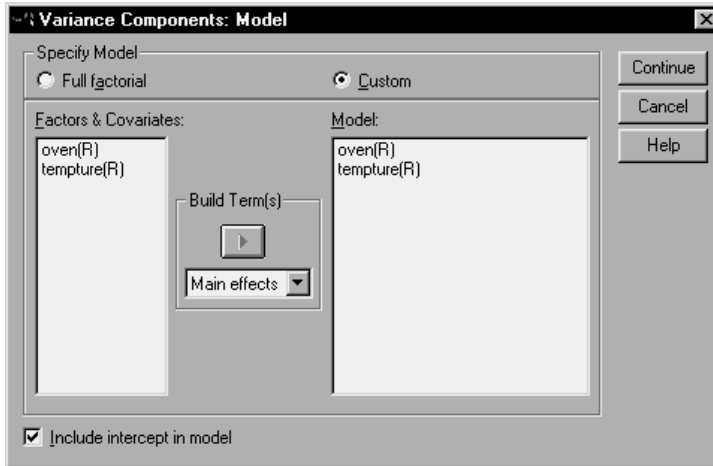
Figure 3-1
Variance Components dialog box



- ▶ Select a dependent variable.
- ▶ Select variables for Fixed Factor(s), Random Factor(s), and Covariate(s), as appropriate for your data. For specifying a weight variable, use WLS Weight.

Variance Components Model

Figure 3-2
Variance Components Model dialog box



Specify Model. A full factorial model contains all factor main effects, all covariate main effects, and all factor-by-factor interactions. It does not contain covariate interactions. Select Custom to specify only a subset of interactions or to specify factor-by-covariate interactions. You must indicate all of the terms to be included in the model.

Factors and Covariates. The factors and covariates are listed with (F) for a fixed factor, (R) for a random factor, and (C) for a covariate.

Model. The model depends on the nature of your data. After selecting Custom, you can select the main effects and interactions that are of interest in your analysis. The model must contain a random factor.

Include intercept in model. Usually the intercept is included in the model. If you can assume that the data pass through the origin, you can exclude the intercept.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

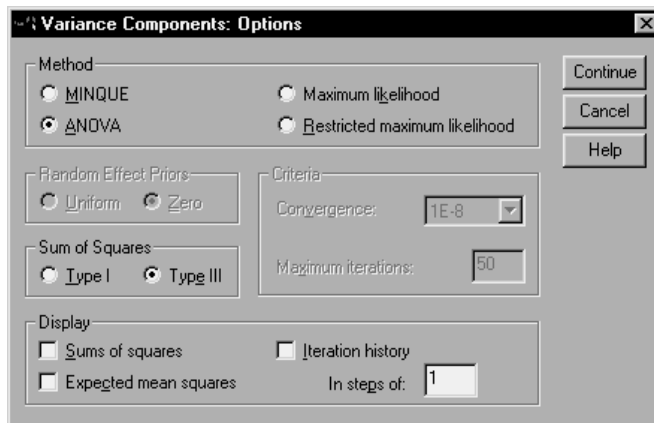
All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Variance Components Options

Figure 3-3
Variance Components Options dialog box



Method. You can choose one of four methods to estimate the variance components.

- **MINQUE** (minimum norm quadratic unbiased estimator) produces estimates that are invariant with respect to the fixed effects. If the data are normally distributed and the estimates are correct, this method produces the least variance among all unbiased estimators. You can choose a method for random-effect prior weights.

- ANOVA (analysis of variance) computes unbiased estimates using either the Type I or Type III sums of squares for each effect. The ANOVA method sometimes produces negative variance estimates, which can indicate an incorrect model, an inappropriate estimation method, or a need for more data.
- Maximum likelihood (ML) produces estimates that would be most consistent with the data actually observed, using iterations. These estimates can be biased. This method is asymptotically normal. ML and REML estimates are invariant under translation. This method does not take into account the degrees of freedom used to estimate the fixed effects.
- Restricted maximum likelihood (REML) estimates reduce the ANOVA estimates for many (if not all) cases of balanced data. Because this method is adjusted for the fixed effects, it should have smaller standard errors than the ML method. This method takes into account the degrees of freedom used to estimate the fixed effects.

Random-Effect Priors. Uniform implies that all random effects and the residual term have an equal impact on the observations. The Zero scheme is equivalent to assuming zero random-effect variances. Available only for the MINQUE method.

Sum of Squares. Type I sums of squares are used for the hierarchical model, which is often used in variance component literature. If you choose Type III, the default in GLM, the variance estimates can be used in GLM Univariate for hypothesis testing with Type III sums of squares. Available only for the ANOVA method.

Criteria. You can specify the convergence criterion and the maximum number of iterations. Available only for the ML or REML methods.

Display. For the ANOVA method, you can choose to display sums of squares and expected mean squares. If you selected Maximum likelihood or Restricted maximum likelihood, you can display a history of the iterations.

Sum of Squares (Variance Components)

For the model, you can choose a type of sum of squares. Type III is the most commonly used and is the default.

Type I. This method is also known as the hierarchical decomposition of the sum-of-squares method. Each term is adjusted for only the term that precedes it in the model. The Type I sum-of-squares method is commonly used for:

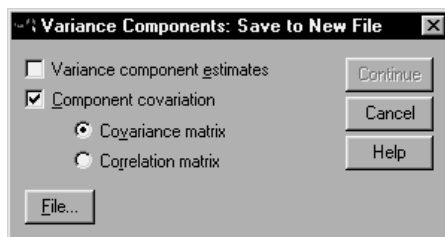
- A balanced ANOVA model in which any main effects are specified before any first-order interaction effects, any first-order interaction effects are specified before any second-order interaction effects, and so on.
- A polynomial regression model in which any lower-order terms are specified before any higher-order terms.
- A purely nested model in which the first-specified effect is nested within the second-specified effect, the second-specified effect is nested within the third, and so on. (This form of nesting can be specified only by using syntax.)

Type III. This method, the default, calculates the sums of squares of an effect in the design as the sums of squares adjusted for any other effects that do not contain it and orthogonal to any effects (if any) that contain it. The Type III sums of squares have one major advantage in that they are invariant with respect to the cell frequencies as long as the general form of estimability remains constant. Therefore, this type is often considered useful for an unbalanced model with no missing cells. In a factorial design with no missing cells, this method is equivalent to the Yates' weighted-squares-of-means technique. The Type III sum-of-squares method is commonly used for:

- Any models listed in Type I.
- Any balanced or unbalanced model with no empty cells.

Variance Components Save to a File

Figure 3-4
Variance Components Save to New File dialog box



You can save some results of this procedure to a new SPSS data file.

Variance component estimates. Saves estimates of the variance components and estimate labels to a data file. These can be used in calculating more statistics or in further analysis in the GLM procedures. For example, you can use them to calculate confidence intervals or test hypotheses.

Component covariation. Saves a variance-covariance matrix or a correlation matrix to a data file. Available only if Maximum likelihood or Restricted maximum likelihood has been specified.

Save to File. Allows you to specify a filename for the file containing the variance component estimates and/or the matrix.

You can use the MATRIX command to extract the data you need from the new file and then compute confidence intervals or perform tests.

VARCOMP Command Additional Features

The SPSS command language also allows you to:

- Specify nested effects in the design (using the DESIGN subcommand).
- Include user-missing values (using the MISSING subcommand).
- Specify EPS criteria (using the CRITERIA subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

Linear Mixed Models

The Linear Mixed Models procedure expands the general linear model so that the data are permitted to exhibit correlated and non-constant variability. The mixed linear model, therefore, provides the flexibility of modeling not only the means of the data but their variances and covariances as well.

The Linear Mixed Models procedure is also a flexible tool for fitting other models that can be formulated as mixed linear models. Such models include multilevel models, hierarchical linear models, and random coefficient models.

Example. A grocery store chain is interested in the effects of various coupons on customer spending. Taking a random sample of their regular customers, they follow the spending of each customer for 10 weeks. In each week, a different coupon is mailed to the customers. Linear Mixed Models is used to estimate the effect of different coupons on spending while adjusting for correlation due to repeated observations on each subject over the 10 weeks.

Methods. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimation.

Statistics. Descriptive statistics: sample sizes, means, and standard deviations of the dependent variable and covariates for each distinct level combination of the factors. Factor-level information: sorted values of the levels of each factor and their frequencies. Also, parameter estimates and confidence intervals for fixed effects, Wald tests and confidence intervals for parameters of covariance matrices. Type I and Type III sums of squares can be used to evaluate different hypotheses. Type III is the default.

Linear Mixed Models Data Considerations

Data. The dependent variable should be quantitative. Factors should be categorical and can have numeric values or string values. Covariates and the weight variable should be quantitative. Subjects and repeated variables may be of any type.

Assumptions. The dependent variable is assumed to be linearly related to the fixed factors, random factors, and covariates. The fixed effects model the mean of the dependent variable. The random effects model the covariance structure of the dependent variable. Multiple random effects are considered independent of each other, and separate covariance matrices will be computed for each; however, model terms specified on the same random effect can be correlated. The repeated measures model the covariance structure of the residuals. The dependent variable is also assumed to come from a normal distribution.

Related procedures. Use the Explore procedure to examine the data before running an analysis. If you do not suspect there to be correlated or nonconstant variability, you can alternatively use the GLM Univariate or GLM Repeated Measures procedure. You can alternatively use the Variance Components Analysis procedure if the random effects have a variance components covariance structure and there are no repeated measures.

Linear Mixed Models Select Subjects/Repeated Variables

This dialog box allows you to select variables that define subjects and repeated observations and to choose a covariance structure for the residuals.

Subjects. A subject is an observational unit that can be considered independent of other subjects. For example, the blood pressure readings from a patient in a medical study can be considered independent of the readings from other patients. Defining subjects becomes particularly important when there are repeated measurements per subject and you want to model the correlation between these observations. For example, you might expect that blood pressure readings from a single patient during consecutive visits to the doctor are correlated.

Subjects can also be defined by the factor-level combination of multiple variables; for example, you can specify *Gender* and *Age category* as subject variables to model the belief that *males over the age of 65* are similar to each other but independent of *males under 65* and *females*.

All of the variables specified in the Subjects list are used to define subjects for the residual covariance structure. You can use some or all of the variables to define subjects for the random-effects covariance structure.

Repeated. The variables specified in this list are used to identify repeated observations. For example, a single variable *Week* might identify the 10 weeks of observations in a medical study, or *Month* and *Day* might be used together to identify daily observations over the course of a year.

Repeated Covariance type. This specifies the covariance structure for the residuals. The available structures are as follows:

- Ante-Dependence: First Order
- AR(1)
- AR(1): Heterogeneous
- ARMA(1,1)
- Compound Symmetry
- Compound Symmetry: Correlation Metric
- Compound Symmetry: Heterogeneous
- Diagonal
- Factor Analytic: First Order
- Factor Analytic: First Order, Heterogeneous
- Huynh-Feldt
- Scaled Identity
- Toeplitz
- Toeplitz: Heterogeneous
- Unstructured
- Unstructured: Correlations

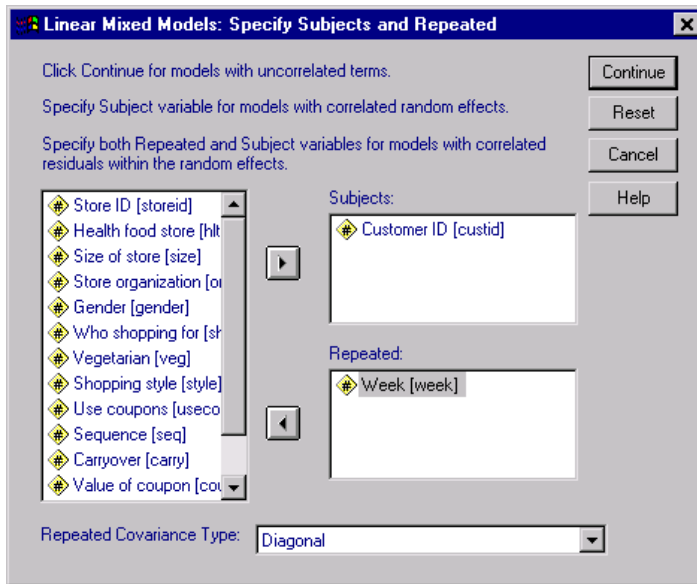
For more information, see “Covariance Structures” in Appendix B on page 125.

Selecting Subjects/Repeated Variables for Linear Mixed Models

- ▶ From the menus choose:
Analyze
Mixed Models
Linear...

Figure 4-1

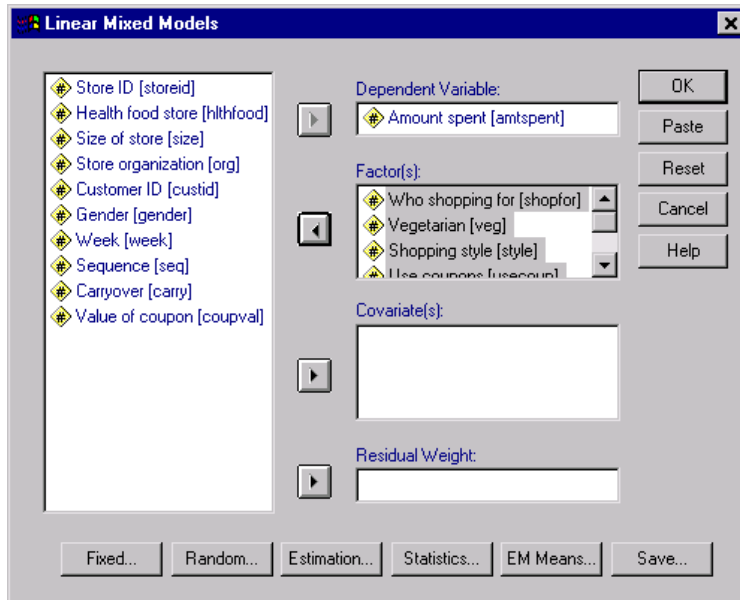
Linear Mixed Models Specify Subjects/Repeated Variables dialog box



- ▶ Optionally, select one or more subjects variables.
- ▶ Optionally, select one or more repeated variables.
- ▶ Optionally, select a residual covariance structure.
- ▶ Click Continue.

Obtaining a Linear Mixed Models Analysis

Figure 4-2
Linear Mixed Models dialog box

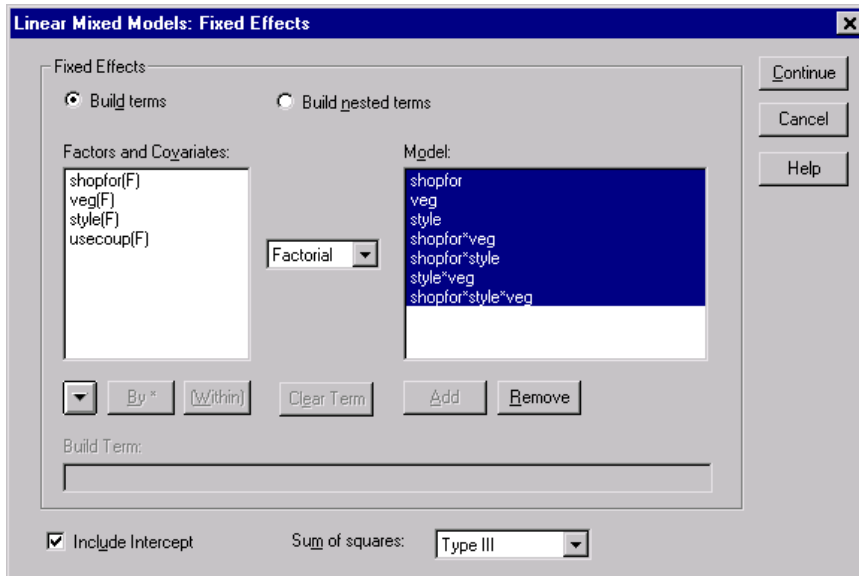


- ▶ Select a dependent variable.
- ▶ Select at least one factor or covariate.
- ▶ Click Fixed or Random and specify at least a fixed-effects or random-effects model.

Optionally, select a weighting variable.

Linear Mixed Models Fixed Effects

Figure 4-3
Linear Mixed Models Fixed Effects dialog box



Fixed Effects. There is no default model, so you must explicitly specify the fixed effects. Alternatively, you can build nested or non-nested terms.

Include Intercept. The intercept is usually included in the model. If you can assume the data pass through the origin, you can exclude the intercept.

Sum of squares. The method of calculating the sums of squares. For models with no missing cells, the Type III method is most commonly used.

Build Non-Nested Terms

For the selected factors and covariates:

Factorial. Creates all possible interactions and main effects of the selected variables. This is the default.

Interaction. Creates the highest-level interaction term of all selected variables.

Main Effects. Creates a main-effects term for each variable selected.

All 2-Way. Creates all possible two-way interactions of the selected variables.

All 3-Way. Creates all possible three-way interactions of the selected variables.

All 4-Way. Creates all possible four-way interactions of the selected variables.

All 5-Way. Creates all possible five-way interactions of the selected variables.

Build Nested Terms

You can build nested terms for your model in this procedure. Nested terms are useful for modeling the effect of a factor or covariate whose values do not interact with the levels of another factor. For example, a grocery store chain may follow the spending of their customers at several store locations. Since each customer frequents only one of those locations, the *Customer* effect can be said to be **nested within** the *Store location* effect.

Additionally, you can include interaction effects or add multiple levels of nesting to the nested term.

Limitations. Nested terms have the following restrictions:

- All factors within an interaction must be unique. Thus, if A is a factor, then specifying $A*A$ is invalid.
- All factors within a nested effect must be unique. Thus, if A is a factor, then specifying $A(A)$ is invalid.
- No effect can be nested within a covariate. Thus, if A is a factor and X is a covariate, then specifying $A(X)$ is invalid.

Sum of Squares

For the model, you can choose a type of sums of squares. Type III is the most commonly used and is the default.

Type I. This method is also known as the hierarchical decomposition of the sum-of-squares method. Each term is adjusted only for the term that precedes it in the model. Type I sums of squares are commonly used for:

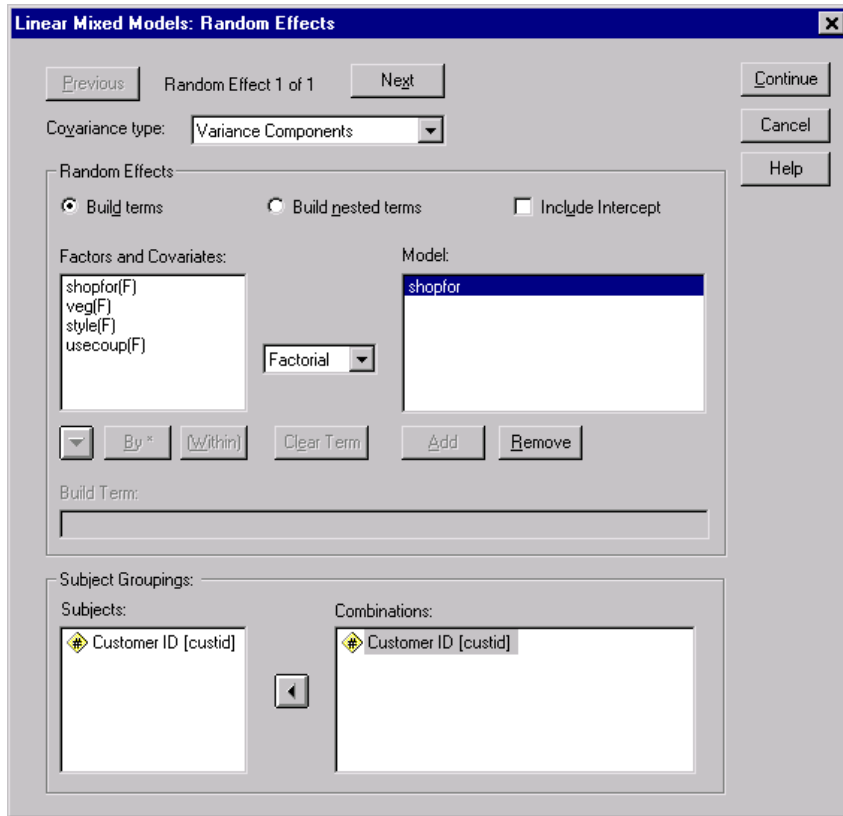
- A balanced ANOVA model in which any main effects are specified before any first-order interaction effects, any first-order interaction effects are specified before any second-order interaction effects, and so on.
- A polynomial regression model in which any lower-order terms are specified before any higher-order terms.
- A purely nested model in which the first-specified effect is nested within the second-specified effect, the second-specified effect is nested within the third, and so on. (This form of nesting can be specified only by using syntax.)

Type III. The default. This method calculates the sums of squares of an effect in the design as the sums of squares adjusted for any other effects that do not contain it and orthogonal to any effects (if any) that contain it. The Type III sums of squares have one major advantage in that they are invariant with respect to the cell frequencies as long as the general form of estimability remains constant. Hence, this type of sums of squares is often considered useful for an unbalanced model with no missing cells. In a factorial design with no missing cells, this method is equivalent to the Yates' weighted-squares-of-means technique. The Type III sum-of-squares method is commonly used for:

- Any models listed in Type I.
- Any balanced or unbalanced model with no empty cells.

Linear Mixed Models Random Effects

Figure 4-4
Linear Mixed Models Random Effects dialog box



Random Effects. There is no default model, so you must explicitly specify the random effects. Alternatively, you can build nested or non-nested terms. You can also choose to include an intercept term in the random-effects model.

You can specify multiple random-effects models. After building the first model, click Next to build the next model. Click Previous to scroll back through existing models. Each random-effect model is assumed to be independent of every other random-effect model; that is, separate covariance matrices will be computed for each. Terms specified in the same random-effect model can be correlated.

Covariance type. This allows you to specify the covariance structure for the random-effects model. A separate covariance matrix is estimated for each random effect. The available structures are as follows:

- Ante-Dependence: First Order
- AR(1)
- AR(1): Heterogeneous
- ARMA(1,1)
- Compound Symmetry
- Compound Symmetry: Correlation Metric
- Compound Symmetry: Heterogeneous
- Diagonal
- Factor Analytic: First Order
- Factor Analytic: First Order, Heterogeneous
- Huynh-Feldt
- Scaled Identity
- Toeplitz
- Toeplitz: Heterogeneous
- Unstructured
- Unstructured: Correlation Metric
- Variance Components

For more information, see “Covariance Structures” in Appendix B on page 125.

Subject Groupings. The variables listed are those that you selected as subject variables in the Select Subjects/Repeated Variables dialog box. Choose some or all of these in order to define the subjects for the random-effects model.

Linear Mixed Models Estimation

Figure 4-5
Linear Mixed Models Estimation dialog box

Linear Mixed Models: Estimation

Method

Restricted Maximum Likelihood (REML)

Maximum Likelihood (ML)

Iterations

Maximum iterations: 100

Maximum step-halvings: 5

Print iteration history for every 1 step(s)

Log-likelihood Convergence

Absolute Relative

Value: 0

Parameter Convergence

Absolute Relative

Value: 0.000001

Hessian Convergence

Absolute Relative

Value: 0

Maximum scoring steps: 1

Singularity tolerance: 0.000000000001

Continue

Cancel

Help

Method. Select the maximum likelihood or restricted maximum likelihood estimation.

Iterations:

- **Maximum iterations.** Specify a non-negative integer.

- **Maximum step-halvings.** At each iteration, the step size is reduced by a factor of 0.5 until the log-likelihood increases or maximum step-halving is reached. Specify a positive integer.
- **Print iteration history for every n step(s).** Displays a table containing the log-likelihood function value and parameter estimates at every n iterations beginning with the 0th iteration (the initial estimates). If you choose to print the iteration history, the last iteration is always printed regardless of the value of n .

Log-likelihood Convergence. Convergence is assumed if the absolute change or relative change in the log-likelihood function is less than the value specified, which must be non-negative. The criterion is not used if the value specified equals 0.

Parameter Convergence. Convergence is assumed if the maximum absolute change or maximum relative change in the parameter estimates is less than the value specified, which must be non-negative. The criterion is not used if the value specified equals 0.

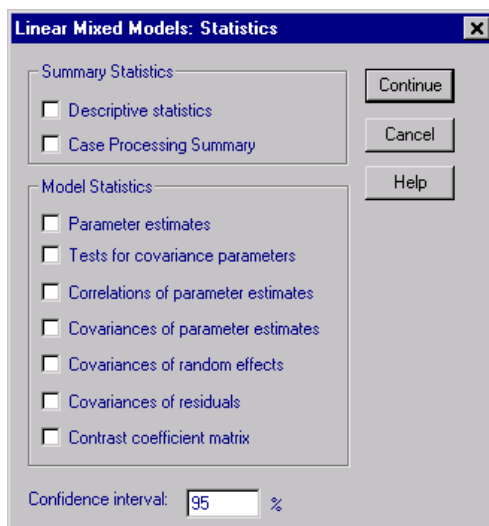
Hessian Convergence. For the Absolute specification, convergence is assumed if a statistic based on the Hessian is less than the value specified. For the Relative specification, convergence is assumed if the statistic is less than the product of the value specified and the absolute value of the log-likelihood. The criterion is not used if the value specified equals 0.

Maximum scoring steps. Requests to use the Fisher scoring algorithm up to iteration number n . Specify a positive integer.

Singularity tolerance. This is the value used as tolerance in checking singularity. Specify a positive value.

Linear Mixed Models Statistics

Figure 4-6
Linear Mixed Models Statistics dialog box



Summary Statistics. Produces tables for:

- **Descriptive statistics.** Displays the sample sizes, means, and standard deviations of the dependent variable and covariates (if specified). These statistics are displayed for each distinct level combination of the factors.
- **Case Processing Summary.** Displays the sorted values of the factors, the repeated measure variables, the repeated measure subjects, and the random-effects subjects and their frequencies.

Model Statistics. Produces tables for:

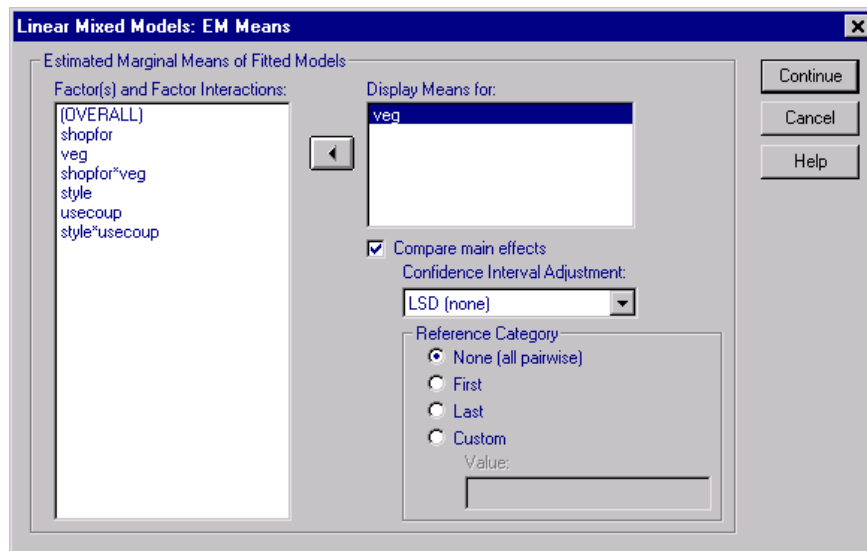
- **Parameter estimates.** Displays the fixed-effects and random-effects parameter estimates and their approximate standard errors.
- **Tests for covariance parameters.** Displays the asymptotic standard errors and Wald tests for the covariance parameters.
- **Correlations of parameter estimates.** Displays the asymptotic correlation matrix of the fixed-effects parameter estimates.

- **Covariances of parameter estimates.** Displays the asymptotic covariance matrix of the fixed-effects parameter estimates.
- **Covariances of random effects.** Displays the estimated covariance matrix of random effects. This option is available only when at least one random effect is specified. If a subject variable is specified for a random effect, then the common block is displayed.
- **Covariances of residuals.** Displays the estimated residual covariance matrix. This option is available only when a repeated variable has been specified. If a subject variable is specified, the common block is displayed.
- **Contrast coefficient matrix.** This option displays the estimable functions used for testing the fixed effects and the custom hypotheses.

Confidence interval. This value is used whenever a confidence interval is constructed. Specify a value greater than or equal to 0 and less than 100. The default value is 95.

Linear Mixed Models EM Means

Figure 4-7
Linear Mixed Models EM Means dialog box

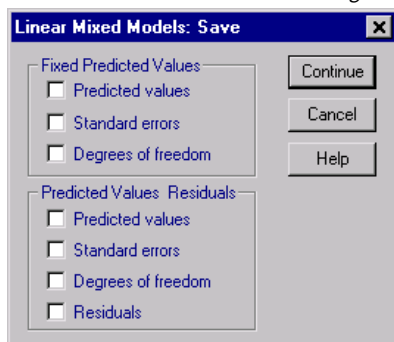


Estimated Marginal Means of Fitted Models. This group allows you to request model-predicted estimated marginal means of the dependent variable in the cells and their standard errors for the specified factors. Moreover, you can request that factor levels of main effects be compared.

- **Factor(s) and Factor Interactions.** This list contains factors and factor interactions that have been specified in the Fixed dialog box, plus an OVERALL term. Model terms built from covariates are excluded from this list.
- **Display Means for.** The procedure will compute the estimated marginal means for factors and factor interactions selected to this list. If OVERALL is selected, the estimated marginal means of the dependent variable are displayed, collapsing over all factors. Note that any selected factors or factor interactions remain selected unless an associated variable has been removed from the Factors list in the main dialog box.
- **Compare main effects.** This option allows you to request pairwise comparisons of levels of selected main effects. The Confidence Interval Adjustment allows you to apply an adjustment to the confidence intervals and significance values to account for multiple comparisons. The available methods are LSD (no adjustment), Bonferroni, and Sidak. Finally, for each factor, you can select a reference category to which comparisons are made. If no reference category is selected, all pairwise comparisons will be constructed. The options for the reference category are first, last, or custom-specified (in which case, you enter the value of the reference category).

Linear Mixed Models Save

Figure 4-8
Linear Mixed Models Save dialog box



This dialog box allows you to save various model results to the working file.

Fixed Predicted Values. Saves variables related to the regression means without the effects.

- **Predicted values.** The regression means without the random effects.
- **Standard errors.** The standard errors of the estimates.
- **Degrees of freedom.** The degrees of freedom associated with the estimates.

Predicted Values and Residuals. Saves variables related to the model fitted value.

- **Predicted values.** The model fitted value.
- **Standard errors.** The standard errors of the estimates.
- **Degrees of freedom.** The degrees of freedom associated with the estimates.
- **Residuals.** The data value minus the predicted value.

MIXED Command Additional Features

The SPSS command language also allows you to:

- Specify tests of effects versus a linear combination of effects or a value (using the TEST subcommand).
- Include user-missing values (using the MISSING subcommand).
- Compute estimated marginal means for specified values of covariates (using the WITH keyword of the EMMEANS subcommand).
- Compare simple main effects of interactions (using the EMMEANS subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

Model Selection Loglinear Analysis

The Model Selection Loglinear Analysis procedure analyzes multiway crosstabulations (contingency tables). It fits hierarchical loglinear models to multidimensional crosstabulations using an iterative proportional-fitting algorithm. This procedure helps you find out which categorical variables are associated. To build models, forced entry and backward elimination methods are available. For saturated models, you can request parameter estimates and tests of partial association. A saturated model adds 0.5 to all cells.

Example. In a study of user preference for one of two laundry detergents, researchers counted people in each group, combining various categories of water softness (soft, medium, or hard), previous use of one of the brands, and washing temperature (cold or hot). They found how temperature is related to water softness and also to brand preference.

Statistics. Frequencies, residuals, parameter estimates, standard errors, confidence intervals, and tests of partial association. For custom models, plots of residuals and normal probability plots.

Model Selection Loglinear Analysis Data Considerations

Data. Factor variables are categorical. All variables to be analyzed must be numeric. Categorical string variables can be recoded to numeric variables before starting the model selection analysis.

Avoid specifying many variables with many levels. Such specifications can lead to a situation where many cells have small numbers of observations, and the chi-square values may not be useful.

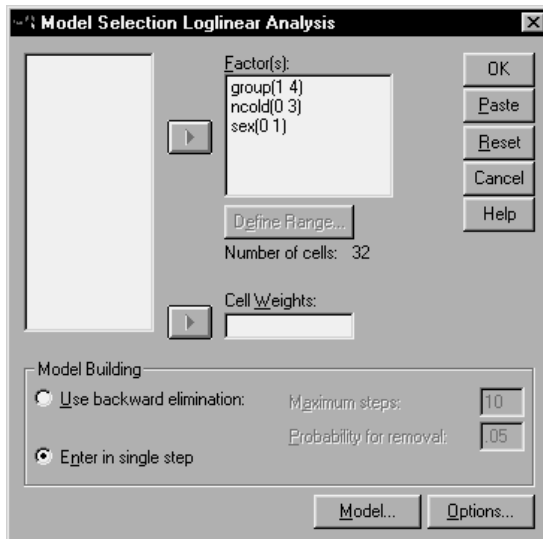
Related procedures. The Model Selection procedure can help identify the terms needed in the model. Then you can continue to evaluate the model using General Loglinear Analysis or Logit Loglinear Analysis. You can use Autorecode to recode string variables. If a numeric variable has empty categories, use Recode to create consecutive integer values.

Obtaining a Model Selection Loglinear Analysis

From the menus choose:

Analyze
Loglinear
Model Selection...

Figure 5-1
Model Selection Loglinear Analysis dialog box

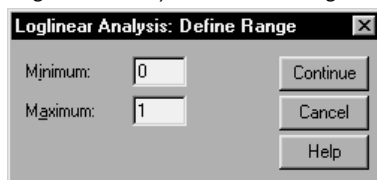


- ▶ Select two or more numeric categorical factors.
- ▶ Select one or more factor variables in the Factor(s) list, and click Define Range.
- ▶ Define the range of values for each factor variable.
- ▶ Select an option in the Model Building group.

Optionally, you can select a cell weight variable to specify structural zeros.

Loglinear Analysis Define Range

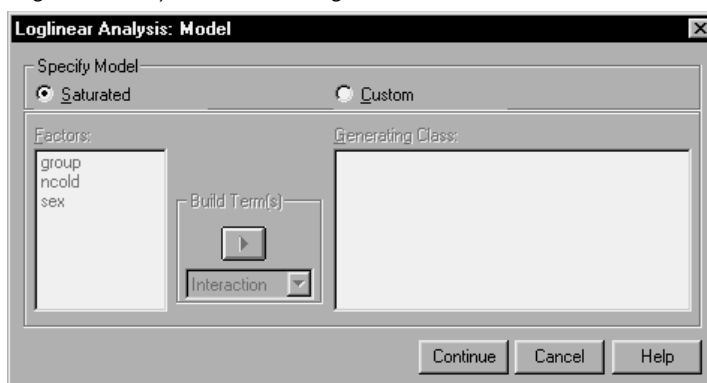
Figure 5-2
Loglinear Analysis Define Range dialog box



You must indicate the range of categories for each factor variable. Values for Minimum and Maximum correspond to the lowest and highest categories of the factor variable. Both values must be integers, and the minimum value must be less than the maximum value. Cases with values outside of the bounds are excluded. For example, if you specify a minimum value of 1 and a maximum value of 3, only the values 1, 2, and 3 are used. Repeat this process for each factor variable.

Loglinear Analysis Model

Figure 5-3
Loglinear Analysis Model dialog box



Specify Model. A saturated model contains all factor main effects and all factor-by-factor interactions. Select Custom to specify a generating class for an unsaturated model.

Generating Class. A generating class is a list of the highest-order terms in which factors appear. SPSS builds a hierarchical model containing the terms that define the generating class and all lower-order relatives. Suppose you select variables *A*, *B*, and *C* in the Factors list and then select Interaction from the Build Terms drop-down list. The resulting model will contain the specified 3-way interaction $A*B*C$, the 2-way interactions $A*B$, $A*C$, and $B*C$, and main effects for *A*, *B*, and *C*. Do not specify the lower-order relatives in the generating class.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

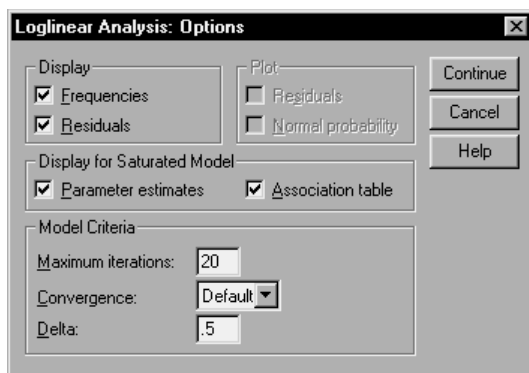
All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Model Selection Loglinear Analysis Options

Figure 5-4
Loglinear Analysis Options dialog box



Display. You can choose Frequencies, Residuals, or both. In a saturated model, the observed and expected frequencies are equal, and the residuals are equal to 0.

Display for Saturated Model. For a saturated model, you can choose Parameter estimates. The parameter estimates may help determine which terms can be dropped from the model. An association table, which lists tests of partial association, is also available. This option is computationally expensive for tables with many factors.

Plot. For custom models, you can choose one or both types of plots, Residuals and Normal probability. These will help determine how well a model fits the data.

Model Criteria. SPSS uses an iterative proportional-fitting algorithm to obtain parameter estimates. You can override one or more of the estimation criteria by specifying Maximum iterations, Convergence, or Delta (a value added to all cell frequencies for saturated models).

HILOGLINEAR Command Additional Features

The SPSS command language also allows you to:

- Specify cell weights in matrix form (using the CWEIGHT subcommand).
- Generate analyses of several models with a single command (using the DESIGN subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

General Loglinear Analysis

The General Loglinear Analysis procedure analyzes the frequency counts of observations falling into each cross-classification category in a crosstabulation or a contingency table. Each cross-classification in the table constitutes a cell, and each categorical variable is called a factor. The dependent variable is the number of cases (frequency) in a cell of the crosstabulation, and the explanatory variables are factors and covariates. This procedure estimates maximum likelihood parameters of hierarchical and nonhierarchical loglinear models using the Newton-Raphson method. Either a Poisson or a multinomial distribution can be analyzed.

You can select up to 10 factors to define the cells of a table. A cell structure variable allows you to define structural zeros for incomplete tables, include an offset term in the model, fit a log-rate model, or implement the method of adjustment of marginal tables. Contrast variables allow computation of generalized log-odds ratios (GLOR).

SPSS automatically displays model information and goodness-of-fit statistics. You can also display a variety of statistics and plots or save residuals and predicted values in the working data file.

Example. Data from a report of automobile accidents in Florida are used to determine the relationship between wearing a seat belt and whether an injury was fatal or nonfatal. The odds ratio indicates significant evidence of a relationship.

Statistics. Observed and expected frequencies; raw, adjusted, and deviance residuals; design matrix; parameter estimates; odds ratio; log-odds ratio; GLOR; Wald statistic; and confidence intervals. Plots: adjusted residuals, deviance residuals, and normal probability.

General Loglinear Analysis Data Considerations

Data. Factors are categorical, and cell covariates are continuous. When a covariate is in the model, SPSS applies the mean covariate value for cases in a cell to that cell. Contrast variables are continuous. They are used to compute generalized log-odds ratios. The values of the contrast variable are the coefficients for the linear combination of the logs of the expected cell counts.

A cell structure variable assigns weights. For example, if some of the cells are structural zeros, the cell structure variable has a value of either 0 or 1. Do not use a cell structure variable to weight aggregated data. Instead, choose Weight Cases from the Data menu.

Assumptions. Two distributions are available in General Loglinear Analysis: Poisson and multinomial.

Under the Poisson distribution assumption:

- The total sample size is not fixed before the study, or the analysis is not conditional on the total sample size.
- The event of an observation being in a cell is statistically independent of the cell counts of other cells.

Under the multinomial distribution assumption:

- The total sample size is fixed, or the analysis is conditional on the total sample size.
- The cell counts are not statistically independent.

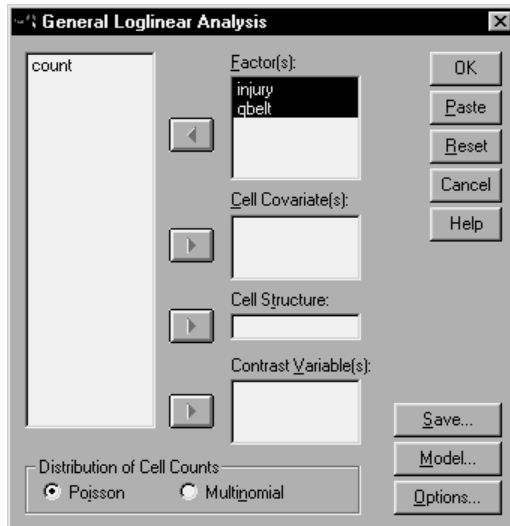
Related procedures. Use the Crosstabs procedure to examine the crosstabulations. Use the Logit Loglinear procedure when it is natural to regard one or more categorical variables as the response variables and the others as the explanatory variables.

Obtaining a General Loglinear Analysis

- ▶ From the menus choose:

- Analyze
 - Loglinear
 - General...

Figure 6-1
General Loglinear Analysis dialog box



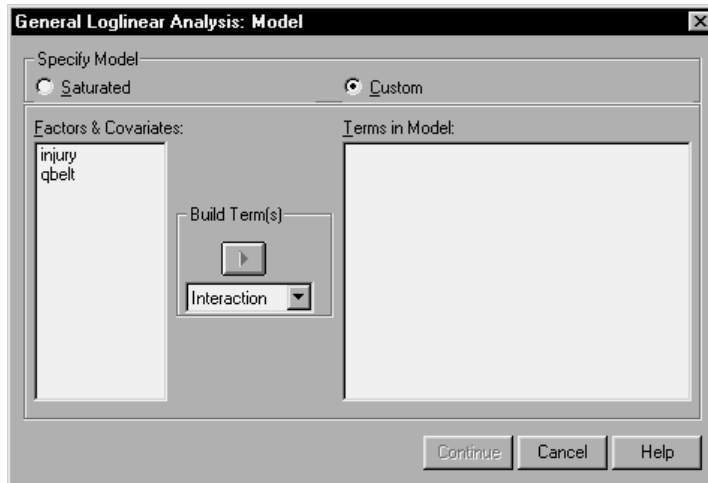
- ▶ In the General Loglinear Analysis dialog box, select up to 10 factor variables.

Optionally, you can:

- Select cell covariates.
- Select a cell structure variable to define structural zeros or include an offset term.
- Select a contrast variable.

General Loglinear Analysis Model

Figure 6-2
General Loglinear Analysis Model dialog box



Specify Model. A saturated model contains all main effects and interactions involving factor variables. It does not contain covariate terms. Select Custom to specify only a subset of interactions or to specify factor-by-covariate interactions.

Factors and Covariates. The factors and covariates are listed, with (Cov) indicating a covariate.

Terms in Model. The model depends on the nature of your data. After selecting Custom, you can select the main effects and interactions that are of interest in your analysis. You must indicate all of the terms to be included in the model.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

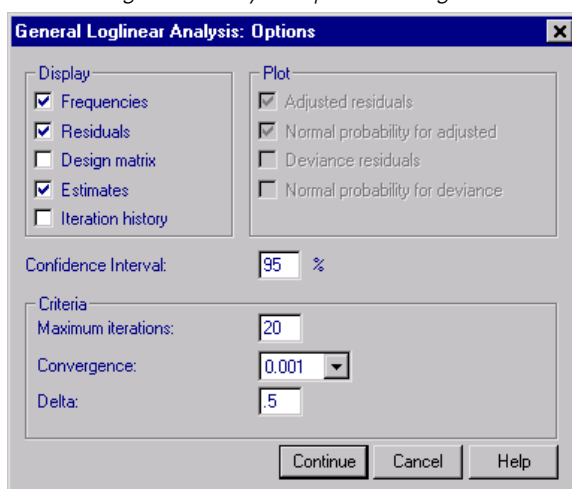
All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

General Loglinear Analysis Options

Figure 6-3
General Loglinear Analysis Options dialog box



The General Loglinear Analysis procedure displays model information and goodness-of-fit statistics. In addition, you can choose one or more of the following:

Display. Several statistics are available for display—observed and expected cell frequencies; raw, adjusted, and deviance residuals; a design matrix of the model; and parameter estimates for the model.

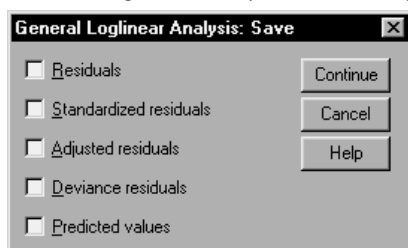
Plot. Plots, available for custom models only, include two scatterplot matrices (adjusted residuals or deviance residuals against observed and expected cell counts). You can also display normal probability and detrended normal plots of adjusted residuals or deviance residuals.

Confidence Interval. The confidence interval for parameter estimates can be adjusted.

Criteria. The Newton-Raphson method is used to obtain maximum likelihood parameter estimates. You can enter new values for the maximum number of iterations, the convergence criterion, and delta (a constant added to all cells for initial approximations). Delta remains in the cells for saturated models.

General Loglinear Analysis Save

Figure 6-4
General Loglinear Analysis Save dialog box



Select the values you want to save as new variables in the working data file. The suffix *n* in the new variable names increments to make a unique name for each saved variable.

The saved values refer to the aggregated data (cells in the contingency table), even if the data are recorded in individual observations in the Data Editor. If you save residuals or predicted values for unaggregated data, the saved value for a cell in the contingency table is entered in the Data Editor for each case in that cell. To make sense of the saved values, you should aggregate the data to obtain the cell counts.

Four types of residuals can be saved: raw, standardized, adjusted, and deviance. The predicted values can also be saved.

- **Residuals.** Also called the simple or raw residual, it is the difference between the observed cell count and its expected count.
- **Standardized Residuals.** The residual divided by an estimate of its standard error. Standardized residuals are also known as Pearson residuals.

- **Adjusted Residuals.** The standardized residual divided by its estimated standard error. Since the adjusted residuals are asymptotically standard normal when the selected model is correct, they are preferred over the standardized residuals for checking for normality.
- **Deviance Residuals.** The signed square root of an individual contribution to the likelihood-ratio chi-square statistic (G^2), where the sign is the sign of the residual (observed count minus expected count). Deviance residuals have an asymptotic standard normal distribution.

GENLOG Command Additional Features

The SPSS command language also allows you to:

- Calculate linear combinations of observed cell frequencies and expected cell frequencies, and print residuals, standardized residuals, and adjusted residuals of that combination (using the GERESID subcommand).
- Change the default threshold value for redundancy checking (using the CRITERIA subcommand).
- Display the standardized residuals (using the PRINT subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

Logit Loglinear Analysis

The Logit Loglinear Analysis procedure analyzes the relationship between dependent (or response) variables and independent (or explanatory) variables. The dependent variables are always categorical, while the independent variables can be categorical (factors). Other independent variables, cell covariates, can be continuous, but they are not applied on a case-by-case basis. The weighted covariate mean for a cell is applied to that cell. The logarithm of the odds of the dependent variables is expressed as a linear combination of parameters. A multinomial distribution is automatically assumed; these models are sometimes called multinomial logit models. This procedure estimates parameters of logit loglinear models using the Newton-Raphson algorithm.

You can select from 1 to 10 dependent and factor variables combined. A cell structure variable allows you to define structural zeros for incomplete tables, include an offset term in the model, fit a log-rate model, or implement the method of adjustment of marginal tables. Contrast variables allow computation of generalized log-odds ratios (GLOR). The values of the contrast variable are the coefficients for the linear combination of the logs of the expected cell counts.

SPSS automatically displays model information and goodness-of-fit statistics. You can also display a variety of statistics and plots or save residuals and predicted values in the working data file.

Example. A study in Florida included 219 alligators. How does the alligators' food type vary with their size and the four lakes in which they live? The study found that the odds of a smaller alligator preferring reptiles to fish is 0.70 times lower than for larger alligators; also, the odds of selecting primarily reptiles instead of fish were highest in lake 3.

Statistics. Observed and expected frequencies; raw, adjusted, and deviance residuals; design matrix; parameter estimates; generalized log-odds ratio; Wald statistic; and confidence intervals. Plots: adjusted residuals, deviance residuals, and normal probability plots.

Logit Loglinear Analysis Data Considerations

Data. The dependent variables are categorical. Factors are categorical. Cell covariates can be continuous, but when a covariate is in the model, SPSS applies the mean covariate value for cases in a cell to that cell. Contrast variables are continuous. They are used to compute generalized log-odds ratios (GLOR). The values of the contrast variable are the coefficients for the linear combination of the logs of the expected cell counts.

A cell structure variable assigns weights. For example, if some of the cells are structural zeros, the cell structure variable has a value of either 0 or 1. Do not use a cell structure variable to weight aggregate data. Instead, use Weight Cases on the Data menu.

Assumptions. The counts within each combination of categories of explanatory variables are assumed to have a multinomial distribution. Under the multinomial distribution assumption:

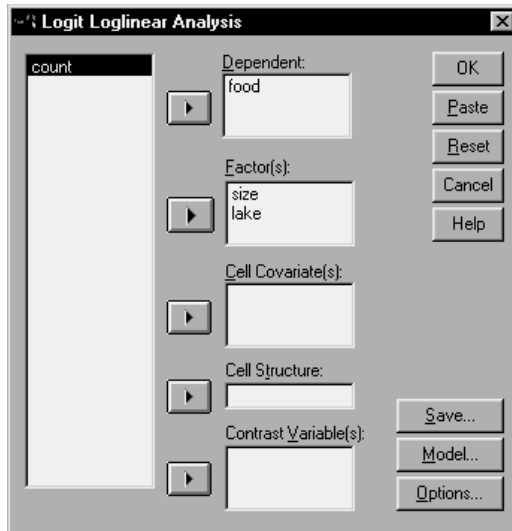
- The total sample size is fixed, or the analysis is conditional on the total sample size.
- The cell counts are not statistically independent.

Related procedures. Use the Crosstabs procedure to display the contingency tables. Use the General Loglinear Analysis procedure when you want to analyze the relationship between an observed count and a set of explanatory variables.

Obtaining a Logit Loglinear Analysis

- ▶ From the menus choose:
 - Analyze
 - Loglinear
 - Logit...

Figure 7-1
Logit Loglinear Analysis dialog box



- ▶ In the Logit Loglinear Analysis dialog box, select one or more dependent variables.
- ▶ Select one or more factor variables.

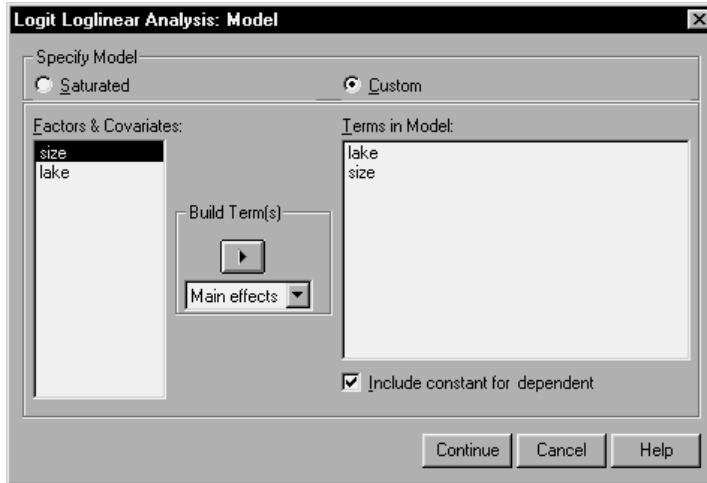
The total number of dependent and factor variables must be less than or equal to 10.

Optionally, you can:

- Select cell covariates.
- Select a cell structure variable to define structural zeros or include an offset term.
- Select one or more contrast variables.

Logit Loglinear Analysis Model

Figure 7-2
Logit Loglinear Analysis Model dialog box



Specify Model. A saturated model contains all main effects and interactions involving factor variables. It does not contain covariate terms. Select Custom to specify only a subset of interactions or to specify factor-by-covariate interactions.

Factors and Covariates. The factors and covariates are listed, with (Cov) indicating a covariate.

Terms in Model. The model depends on the nature of your data. After selecting Custom, you can select the main effects and interactions that are of interest in your analysis. You must indicate all of the terms to be included in the model.

Terms are added to the design by taking all possible combinations of the dependent terms and matching each combination with each term in the model list. If Include constant for dependent is selected, there is also a unit term (1) added to the model list.

For example, suppose variables $D1$ and $D2$ are the dependent variables. A dependent terms list is created by the Logit Loglinear Analysis procedure ($D1$, $D2$, $D1*D2$). If the Terms in Model list contains $M1$ and $M2$ and a constant is included, the model list contains 1, $M1$, and $M2$. The resultant design includes combinations of each model term with each dependent term:

$D1$, $D2$, $D1*D2$

$M1 * D1, M1 * D2, M1 * D1 * D2$

$M2 * D1, M2 * D2, M2 * D1 * D2$

Include constant for dependent. Includes a constant for the dependent variable in a custom model.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables. This is the default.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

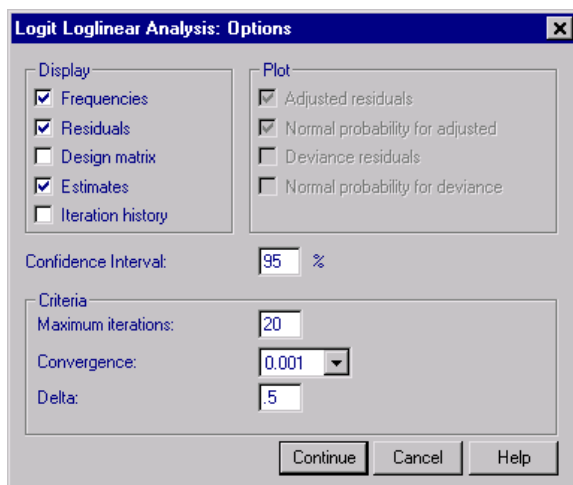
All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Logit Loglinear Analysis Options

Figure 7-3
Logit Loglinear Analysis Options dialog box



The Logit Loglinear Analysis procedure displays model information and goodness-of-fit statistics. In addition, you can choose one or more of the following options:

Display. Several statistics are available for display: observed and expected cell frequencies; raw, adjusted, and deviance residuals; a design matrix of the model; and parameter estimates for the model.

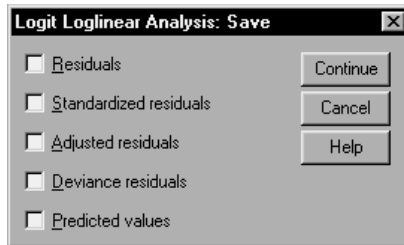
Plot. Plots available for custom models include two scatterplot matrices (adjusted residuals or deviance residuals against observed and expected cell counts). You can also display normal probability and detrended normal plots of adjusted residuals or deviance residuals.

Confidence Interval. The confidence interval for parameter estimates can be adjusted.

Criteria. The Newton-Raphson method is used to obtain maximum likelihood parameter estimates. You can enter new values for the maximum number of iterations, the convergence criterion, and delta (a constant added to all cells for initial approximations). Delta remains in the cells for saturated models.

Logit Loglinear Analysis Save

Figure 7-4
Logit Loglinear Analysis Save dialog box



Select the values you want to save as new variables in the working data file. The suffix *n* in the new variable names increments to make a unique name for each saved variable.

The saved values refer to the aggregated data (to cells in the contingency table), even if the data are recorded in individual observations in the Data Editor. If you save residuals or predicted values for unaggregated data, the saved value for a cell in the contingency table is entered in the Data Editor for each case in that cell. To make sense of the saved values, you should aggregate the data to obtain the cell counts.

Four types of residuals can be saved: raw, standardized, adjusted, and deviance. The predicted values can also be saved.

- **Residuals.** Also called the simple or raw residual, it is the difference between the observed cell count and its expected count.
- **Standardized Residuals.** The residual divided by an estimate of its standard error. Standardized residuals are also known as Pearson residuals.
- **Adjusted Residuals.** The standardized residual divided by its estimated standard error. Since the adjusted residuals are asymptotically standard normal when the selected model is correct, they are preferred over the standardized residuals for checking for normality.
- **Deviance Residuals.** The signed square root of an individual contribution to the likelihood-ratio chi-square statistic (*G* squared), where the sign is the sign of the residual (observed count minus expected count). Deviance residuals have an asymptotic standard normal distribution.

GENLOG Command Additional Features

The SPSS command language also allows you to:

- Calculate linear combinations of observed cell frequencies and expected cell frequencies, and print residuals, standardized residuals, and adjusted residuals of that combination (using the GERESID subcommand).
- Change the default threshold value for redundancy checking (using the CRITERIA subcommand).
- Display the standardized residuals (using the PRINT subcommand).

See the *SPSS Command Syntax Reference* for complete syntax information.

Ordinal Regression

Ordinal Regression allows you to model the dependence of a polytomous ordinal response on a set of predictors, which can be factors or covariates. The design of Ordinal Regression is based on the methodology of McCullagh (1980, 1998), and the procedure is referred to as PLUM in the syntax.

Standard linear regression analysis involves minimizing the sum-of-squared differences between a response (dependent) variable and a weighted combination of predictor (independent) variables. The estimated coefficients reflect how changes in the predictors affect the response. The response is assumed to be numerical, in the sense that changes in the level of the response are equivalent throughout the range of the response. For example, the difference in height between a person who is 150 cm tall and a person who is 140 cm tall is 10 cm, which has the same meaning as the difference in height between a person who is 210 cm tall and a person who is 200 cm tall. These relationships do not necessarily hold for ordinal variables, in which the choice and number of response categories can be quite arbitrary.

Example. Ordinal Regression could be used to study patient reaction to drug dosage. The possible reactions may be classified as “none,” “mild,” “moderate,” or “severe.” The difference between a mild and moderate reaction is difficult or impossible to quantify and is based on perception. Moreover, the difference between a mild and moderate response may be greater or less than the difference between a moderate and severe response.

Statistics and plots. Observed and expected frequencies and cumulative frequencies, Pearson residuals for frequencies and cumulative frequencies, observed and expected probabilities, observed and expected cumulative probabilities of each response category by covariate pattern, asymptotic correlation and covariance matrices of parameter estimates, Pearson's chi-square and likelihood-ratio chi-square, goodness-of-fit statistics, iteration history, test of parallel lines assumption, parameter

estimates, standard errors, confidence intervals, and Cox and Snell's, Nagelkerke's, and McFadden's R^2 statistics.

Ordinal Regression Data Considerations

Data. The dependent variable is assumed to be ordinal and can be numeric or string. The ordering is determined by sorting the values of the dependent variable in ascending order. The lowest value defines the first category. Factor variables are assumed to be categorical. Covariate variables must be numeric. Note that using more than one continuous covariate can easily result in the creation of a very large cell probabilities table.

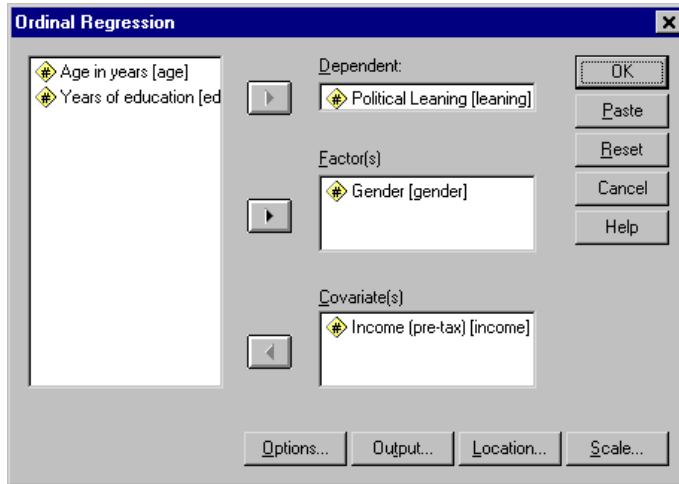
Assumptions. Only one response variable is allowed, and it must be specified. Also, for each distinct pattern of values across the independent variables, the responses are assumed to be independent multinomial variables.

Related procedures. Nominal logistic regression uses similar models for nominal dependent variables.

Obtaining an Ordinal Regression

- ▶ From the menus choose:
 - Analyze
 - Regression
 - Ordinal...

Figure 8-1
Ordinal Regression dialog box

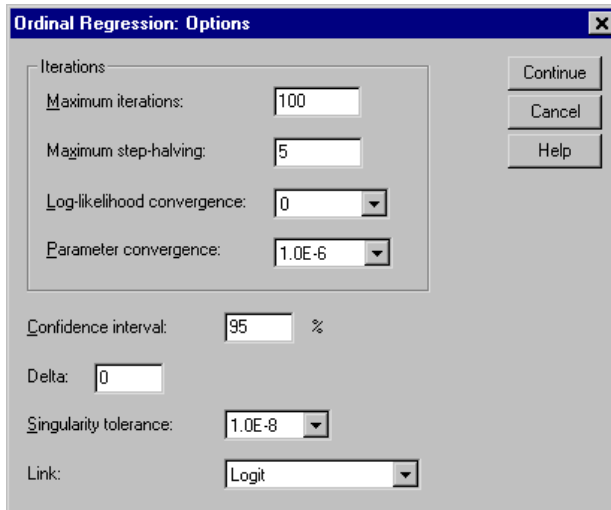


- ▶ Select one dependent variable.
- ▶ Click OK.

Ordinal Regression Options

The Options dialog box allows you to adjust parameters used in the iterative estimation algorithm, choose a level of confidence for your parameter estimates, and select a link function.

Figure 8-2
Ordinal Regression Options dialog box



Iterations. You can customize the iterative algorithm.

- **Maximum iterations.** Specify a non-negative integer. If 0 is specified, the procedure returns the initial estimates.
- **Maximum step-halving.** Specify a positive integer.
- **Log-likelihood convergence.** The algorithm stops if the absolute or relative change in the log-likelihood is less than this value. The criterion is not used if 0 is specified.
- **Parameter convergence.** The algorithm stops if the absolute or relative change in each of the parameter estimates is less than this value. The criterion is not used if 0 is specified.

Confidence interval. Specify a value greater than or equal to 0 and less than 100.

Delta. The value added to zero cell frequencies. Specify a non-negative value less than 1.

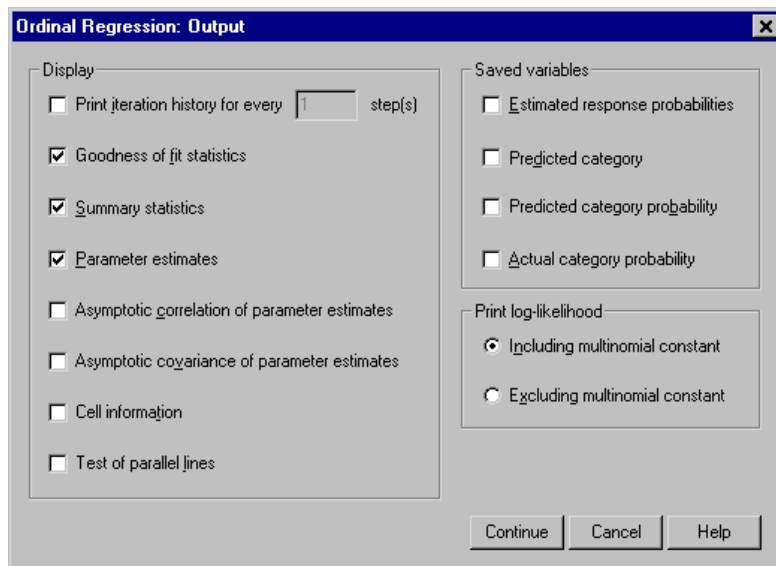
Singularity tolerance. Used for checking for highly dependent predictors. Select a value from the list of options.

Link. Choose among the Cauchit, Complementary Log-log, Logit, Negative Log-log, and Probit functions.

Ordinal Regression Output

The Output dialog box allows you to produce tables for display in the Viewer and save variables to the working file.

Figure 8-3
Ordinal Regression Output dialog box



Display. Produces tables for:

- **Print iteration history.** The log-likelihood and parameter estimates are printed for the print iteration frequency specified. The first and last iterations are always printed.
- **Goodness-of-fit statistics.** The Pearson and likelihood-ratio chi-square statistics. They are computed based on the classification specified in the variable list.
- **Summary statistics.** Cox and Snell's, Nagelkerke's, and McFadden's R^2 statistics.
- **Parameter estimates.** Parameter estimates, standard errors, and confidence intervals.

- **Asymptotic correlation of parameter estimates.** Matrix of parameter estimate correlations.
- **Asymptotic covariance of parameter estimates.** Matrix of parameter estimate covariances.
- **Cell information.** Observed and expected frequencies and cumulative frequencies, Pearson residuals for frequencies and cumulative frequencies, observed and expected probabilities, and observed and expected cumulative probabilities of each response category by covariate pattern. Note that for models with many covariate patterns (for example, models with continuous covariates), this option can generate a very large, unwieldy table.
- **Test of parallel lines.** Test of the hypothesis that the location parameters are equivalent across the levels of the dependent variable. This is available only for the location-only model.

Saved variables. Saves the following variables to the working file:

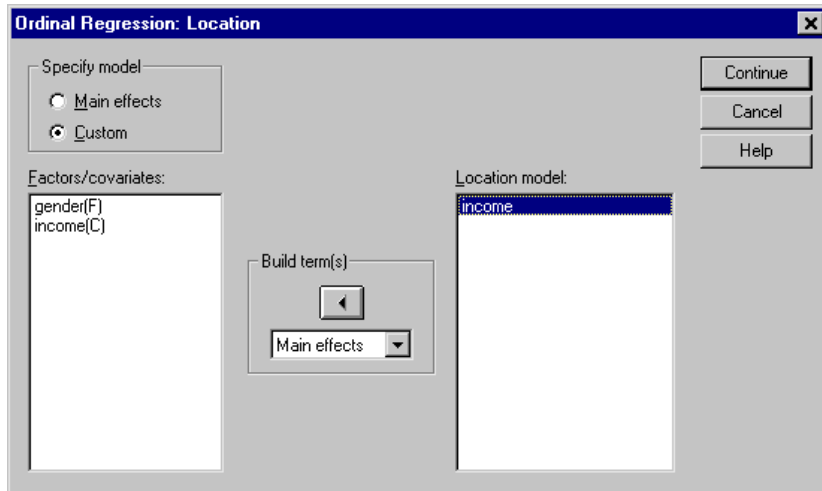
- **Estimated response probabilities.** Model-estimated probabilities of classifying a factor/covariate pattern into the response categories. There are as many probabilities as the number of response categories.
- **Predicted category.** The response category that has the maximum estimated probability for a factor/covariate pattern.
- **Predicted category probability.** Estimated probability of classifying a factor/covariate pattern into the predicted category. This probability is also the maximum of the estimated probabilities of the factor/covariate pattern.
- **Actual category probability.** Estimated probability of classifying a factor/covariate pattern into the actual category.

Print log-likelihood. Controls the display of the log-likelihood. Including multinomial constant gives you the full value of the likelihood. To compare your results across products that do not include the constant, you can choose to exclude it.

Ordinal Regression Location Model

The Location dialog box allows you to specify the location model for your analysis.

Figure 8-4
Ordinal Regression Location dialog box



Specify model. A main-effects model contains the covariate and factor main effects but no interaction effects. You can create a custom model to specify subsets of factor interactions or covariate interactions.

Factors/covariates. The factors and covariates are listed with (F) for factor and (C) for covariate.

Location model. The model depends on the main effects and interaction effects that you select.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

All 3-way. Creates all possible three-way interactions of the selected variables.

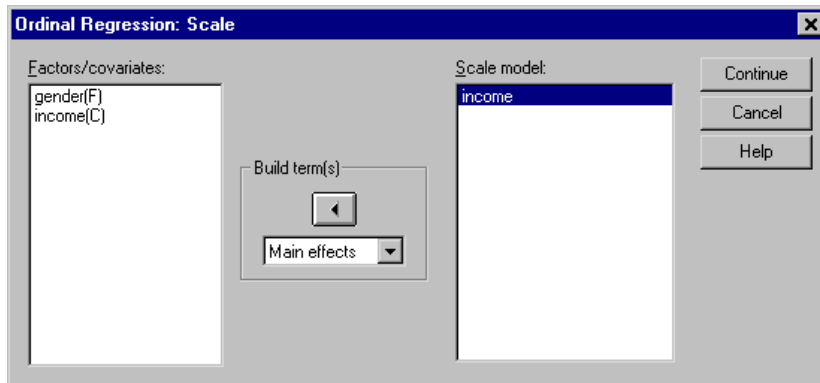
All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

Ordinal Regression Scale Model

The Scale dialog box allows you to specify the scale model for your analysis.

Figure 8-5
Ordinal Regression Scale dialog box



Factors/covariates. The factors and covariates are listed with (F) for factor and (C) for covariate.

Scale model. The model depends on the main and interaction effects that you select.

Build Terms

For the selected factors and covariates:

Interaction. Creates the highest-level interaction term of all selected variables.

Main effects. Creates a main-effects term for each variable selected.

All 2-way. Creates all possible two-way interactions of the selected variables.

All 3-way. Creates all possible three-way interactions of the selected variables.

All 4-way. Creates all possible four-way interactions of the selected variables.

All 5-way. Creates all possible five-way interactions of the selected variables.

PLUM Command Additional Features

You can customize your Ordinal Regression if you paste your selections into a syntax window and edit the resulting PLUM command syntax. The SPSS command language also allows you to:

- Create customized hypothesis tests by specifying null hypotheses as linear combinations of parameters.

See the *SPSS Command Syntax Reference* for complete syntax information.

Life Tables

There are many situations in which you would want to examine the distribution of times between two events, such as length of employment (time between being hired and leaving the company). However, this kind of data usually includes some cases for which the second event isn't recorded (for example, people still working for the company at the end of the study). This can happen for several reasons: for some cases, the event simply doesn't occur before the end of the study; for other cases, we lose track of their status sometime before the end of the study; still other cases may be unable to continue for reasons unrelated to the study (such as an employee becoming ill and taking a leave of absence). Collectively, such cases are known as **censored cases**, and they make this kind of study inappropriate for traditional techniques such as t tests or linear regression.

A statistical technique useful for this type of data is called a follow-up **life table**. The basic idea of the life table is to subdivide the period of observation into smaller time intervals. For each interval, all people who have been observed at least that long are used to calculate the probability of a terminal event occurring in that interval. The probabilities estimated from each of the intervals are then used to estimate the overall probability of the event occurring at different time points.

Example. Is a new nicotine patch therapy better than traditional patch therapy in helping people to quit smoking? You could conduct a study using two groups of smokers, one of which received the traditional therapy and the other of which received the experimental therapy. Constructing life tables from the data would allow you to compare overall abstinence rates between the two groups to determine if the experimental treatment is an improvement over the traditional therapy. You can also plot the survival or hazard functions and compare them visually for more detailed information.

Statistics. Number entering, number leaving, number exposed to risk, number of terminal events, proportion terminating, proportion surviving, cumulative proportion surviving (and standard error), probability density (and standard error), and hazard rate (and standard error) for each time interval for each group; median survival time for each group; and Wilcoxon (Gehan) test for comparing survival distributions between groups. Plots: function plots for survival, log survival, density, hazard rate, and one minus survival.

Life Tables Data Considerations

Data. Your time variable should be quantitative. Your status variable should be dichotomous or categorical, coded as integers, with events being coded as a single value or a range of consecutive values. Factor variables should be categorical, coded as integers.

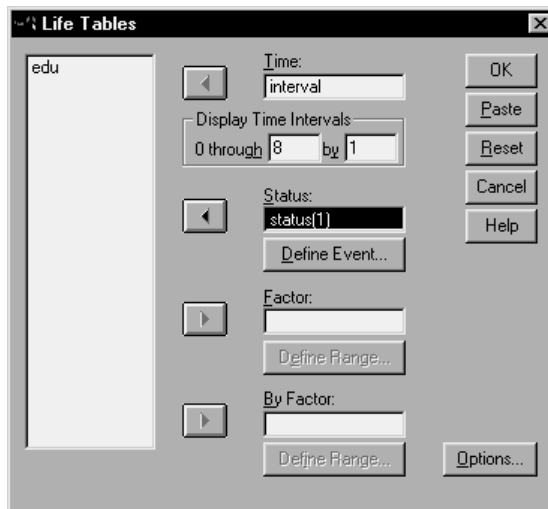
Assumptions. Probabilities for the event of interest should depend only on time after the initial event—they are assumed to be stable with respect to absolute time. That is, cases that enter the study at different times (for example, patients who begin treatment at different times) should behave similarly. There should also be no systematic differences between censored and uncensored cases. If, for example, many of the censored cases are patients with more serious conditions, your results may be biased.

Related procedures. The Life Tables procedure uses an actuarial approach to this kind of analysis (known generally as Survival Analysis). The Kaplan-Meier Survival Analysis procedure uses a slightly different method of calculating life tables that does not rely on partitioning the observation period into smaller time intervals. This method is recommended if you have a small number of observations, such that there would be only a small number of observations in each survival time interval. If you have variables that you suspect are related to survival time or variables that you want to control for (covariates), use the Cox Regression procedure. If your covariates can have different values at different points in time for the same case, use Cox Regression with Time-Dependent Covariates.

Creating Life Tables

- ▶ From the menus choose:
 - Analyze
 - Survival
 - Life Tables...

Figure 9-1
Life Tables dialog box



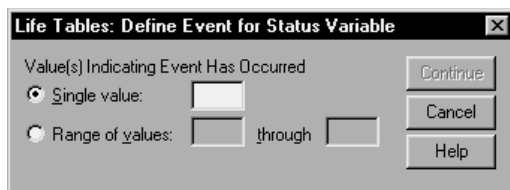
- ▶ Select one **numeric** survival variable.
- ▶ Specify the time intervals to be examined.
- ▶ Select a status variable to define cases for which the terminal event has occurred.
- ▶ Click Define Event to specify the value of the status variable that indicates that an event occurred.

Optionally, you can select a first-order factor variable. Actuarial tables for the survival variable are generated for each category of the factor variable.

You can also select a second-order *by factor* variable. Actuarial tables for the survival variable are generated for every combination of the first- and second-order factor variables.

Life Tables Define Events for Status Variables

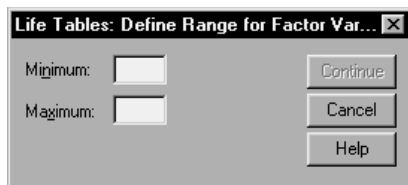
Figure 9-2
Life Tables Define Event for Status Variable dialog box



Occurrences of the selected value or values for the status variable indicate that the terminal event has occurred for those cases. All other cases are considered to be censored. Enter either a single value or a range of values that identifies the event of interest.

Life Tables Define Range

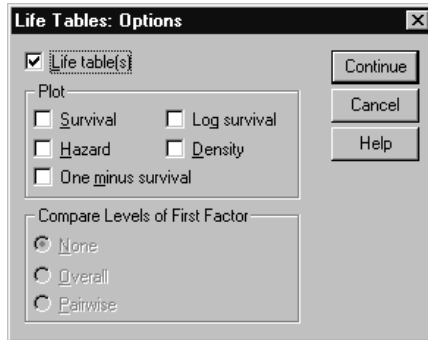
Figure 9-3
Life Tables Define Range dialog box



Cases with values for the factor variable in the range you specify will be included in the analysis, and separate tables (and plots, if requested) will be generated for each unique value in the range.

Life Tables Options

Figure 9-4
Life Tables Options dialog box



You can control various aspects of your Life Tables analysis.

Life tables. To suppress the display of life tables in the output, deselect Life tables.

Plot. Allows you to request plots of the survival functions. If you have defined factor variable(s), plots are generated for each subgroup defined by the factor variable(s). Available plots are survival, log survival, hazard, density, and one minus survival.

- **Survival Plot.** Displays the cumulative survival function on a linear scale.
- **Log Survival Plot.** Displays the cumulative survival function on a logarithmic scale.
- **Hazard Plot.** Displays the cumulative hazard function on a linear scale.
- **Density Plot.** Displays the density function.
- **One Minus Survival.** Plots one-minus the survival function on a linear scale.

Compare Levels of First Factor. If you have a first-order control variable, you can select one of the alternatives in this group to perform the Wilcoxon (Gehan) test, which compares the survival of subgroups. Tests are performed on the first-order factor. If you have defined a second-order factor, tests are performed for each level of the second-order variable.

SURVIVAL Command Additional Features

The SPSS command language also allows you to:

- Specify more than one dependent variable.
- Specify unequally spaced intervals.
- Specify more than one status variable.
- Specify comparisons that do not include all the factor and all the control variables.
- Calculate approximate, rather than exact, comparisons.

See the *SPSS Command Syntax Reference* for complete syntax information.

Kaplan-Meier Survival Analysis

There are many situations in which you would want to examine the distribution of times between two events, such as length of employment (time between being hired and leaving the company). However, this kind of data usually includes some censored cases. Censored cases are cases for which the second event isn't recorded (for example, people still working for the company at the end of the study). The Kaplan-Meier procedure is a method of estimating time-to-event models in the presence of censored cases. The Kaplan-Meier model is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit of those probabilities to estimate the survival rate at each point in time.

Example. Does a new treatment for AIDS have any therapeutic benefit in extending life? You could conduct a study using two groups of AIDS patients, one receiving traditional therapy and the other receiving the experimental treatment. Constructing a Kaplan-Meier model from the data would allow you to compare overall survival rates between the two groups to determine whether the experimental treatment is an improvement over the traditional therapy. You can also plot the survival or hazard functions and compare them visually for more detailed information.

Statistics. Survival table, including time, status, cumulative survival and standard error, cumulative events, and number remaining; and mean and median survival time, with standard error and 95% confidence interval. Plots: survival, hazard, log survival, and one minus survival.

Kaplan-Meier Data Considerations

Data. The time variable should be continuous, the status variable can be categorical or continuous, and the factor and strata variables should be categorical.

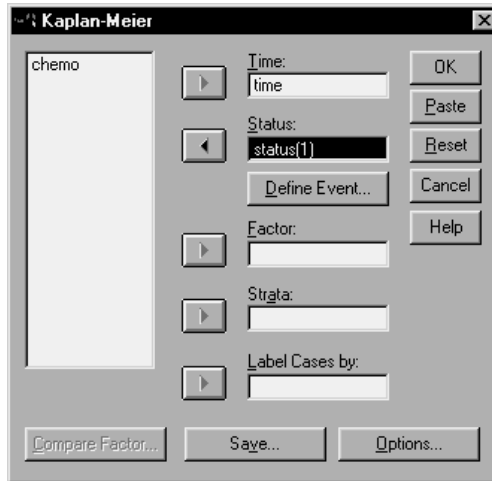
Assumptions. Probabilities for the event of interest should depend only on time after the initial event—they are assumed to be stable with respect to absolute time. That is, cases that enter the study at different times (for example, patients who begin treatment at different times) should behave similarly. There should also be no systematic differences between censored and uncensored cases. If, for example, many of the censored cases are patients with more serious conditions, your results may be biased.

Related procedures. The Kaplan-Meier procedure uses a method of calculating life tables that estimates the survival or hazard function at the time of each event. The Life Tables procedure uses an actuarial approach to survival analysis that relies on partitioning the observation period into smaller time intervals and may be useful for dealing with large samples. If you have variables that you suspect are related to survival time or variables that you want to control for (covariates), use the Cox Regression procedure. If your covariates can have different values at different points in time for the same case, use Cox Regression with Time-Dependent Covariates.

Obtaining a Kaplan-Meier Survival Analysis

- ▶ From the menus choose:
 - Analyze
 - Survival
 - Kaplan-Meier...

Figure 10-1
Kaplan-Meier dialog box

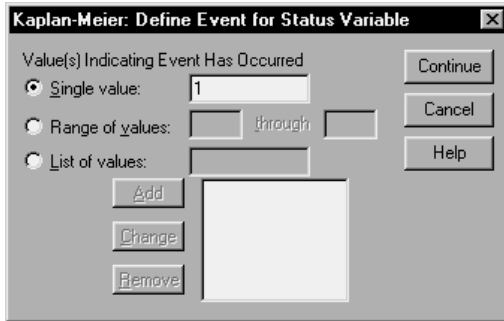


- ▶ Select a time variable.
- ▶ Select a status variable to identify cases for which the terminal event has occurred. This variable can be numeric or **short string**. Then click Define Event.

Optionally, you can select a factor variable to examine group differences. You can also select a strata variable, which will produce separate analyses for each level (stratum) of the variable.

Kaplan-Meier Define Event for Status Variable

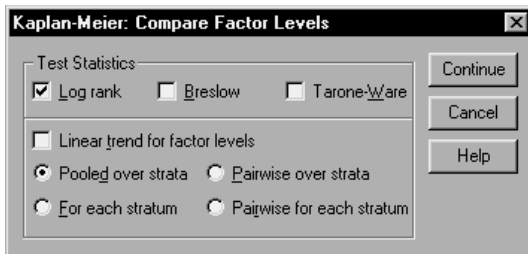
Figure 10-2
Kaplan-Meier Define Event for Status Variable dialog box



Enter the value or values indicating that the terminal event has occurred. You can enter a single value, a range of values, or a list of values. The Range of Values option is available only if your status variable is numeric.

Kaplan-Meier Compare Factor Levels

Figure 10-3
Kaplan-Meier Compare Factor Levels dialog box



You can request statistics to test the equality of the survival distributions for the different levels of the factor. Available statistics are log rank, Breslow, and Tarone-Ware. Select one of the alternatives to specify the comparisons to be made: pooled over strata, for each stratum, pairwise over strata, or pairwise for each stratum.

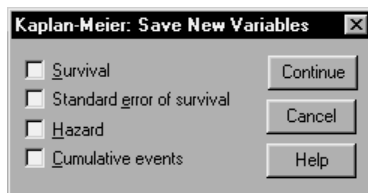
- **Log Rank.** A test for comparing the equality of survival distributions. All time points are weighted equally in this test.

- **Breslow.** A test for comparing the equality of survival distributions. Time points are weighted by the number of cases at risk at each time point.
- **Tarone-Ware.** A test for comparing the equality of survival distributions. Time points are weighted by the square root of the number of cases at risk at each time point.
- **Pooled Over Strata.** Compares all factor levels in a single test to test the equality of survival curves.
- **Pairwise Over Strata.** Compares each distinct pair of factor levels. Pairwise trend tests are not available.
- **For Each Stratum.** Performs a separate test of equality of all factor levels for each stratum. If you do not have a stratification variable, the tests are not performed.
- **Pairwise for each Stratum.** Compares each distinct pair of factor levels for each stratum. Pairwise trend tests are not available. If you do not have a stratification variable, the tests are not performed.

Linear trend for factor levels. Allows you to test for a linear trend across levels of the factor. This option is available only for overall (rather than pairwise) comparisons of factor levels.

Kaplan-Meier Save New Variables

Figure 10-4
Kaplan-Meier Save New Variables dialog box



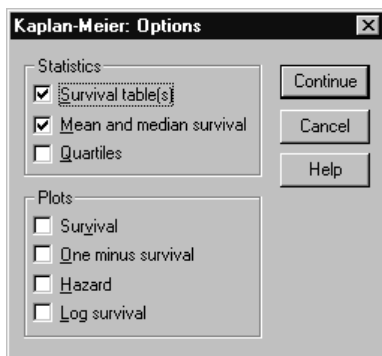
You can save information from your Kaplan-Meier table as new variables, which can then be used in subsequent analyses to test hypotheses or check assumptions. You can save survival, standard error of survival, hazard, and cumulative events as new variables.

- **Survival.** Cumulative survival probability estimate. The default variable name is the prefix `sur_` with a sequential number appended to it. For example, if `sur_1` already exists, Kaplan-Meier assigns the variable name `sur_2`.

- **Standard Error of Survival.** Standard error of the cumulative survival estimate. The default variable name is the prefix `se_` with a sequential number appended to it. For example, if `se_1` already exists, Kaplan-Meier assigns the variable name `se_2`.
- **Hazard.** Cumulative hazard function estimate. The default variable name is the prefix `haz_` with a sequential number appended to it. For example, if `haz_1` already exists, Kaplan-Meier assigns the variable name `haz_2`.
- **Cumulative Events.** Cumulative frequency of events when cases are sorted by their survival times and status codes. The default variable name is the prefix `cum_` with a sequential number appended to it. For example, if `cum_1` already exists, Kaplan-Meier assigns the variable name `cum_2`.

Kaplan-Meier Options

Figure 10-5
Kaplan-Meier Options dialog box



You can request various output types from Kaplan-Meier analysis.

Statistics. You can select statistics displayed for the survival functions computed, including survival table(s), mean and median survival, and quartiles. If you have included factor variables, separate statistics are generated for each group.

Plots. Plots allow you to examine the survival, hazard, log-survival, and one-minus-survival functions visually. If you have included factor variables, functions are plotted for each group.

- **Survival Plot.** Displays the cumulative survival function on a linear scale.
- **One Minus Survival.** Plots one-minus the survival function on a linear scale.

- **Hazard Plot.** Displays the cumulative hazard function on a linear scale.
- **Log Survival Plot.** Displays the cumulative survival function on a logarithmic scale.

KM Command Additional Features

The SPSS command language also allows you to:

- Obtain frequency tables that consider cases lost to follow-up as a separate category from censored cases.
- Specify unequal spacing for the test for linear trend.
- Obtain percentiles other than quartiles for the survival time variable.

See the *SPSS Command Syntax Reference* for complete syntax information.

Cox Regression Analysis

Like Life Tables and Kaplan-Meier survival analysis, Cox Regression is a method for modeling time-to-event data in the presence of censored cases. However, Cox Regression allows you to include predictor variables (covariates) in your models. For example, you could construct a model of length of employment based on educational level and job category. Cox Regression will handle the censored cases correctly, and it will provide estimated coefficients for each of the covariates, allowing you to assess the impact of multiple covariates in the same model. You can also use Cox Regression to examine the effect of continuous covariates.

Example. Do men and women have different risks of developing lung cancer based on cigarette smoking? By constructing a Cox Regression model, with cigarette usage (cigarettes smoked per day) and gender entered as covariates, you can test hypotheses regarding the effects of gender and cigarette usage on time-to-onset for lung cancer.

Statistics. For each model: $-2LL$, the likelihood-ratio statistic, and the overall chi-square. For variables in the model: parameter estimates, standard errors, and Wald statistics. For variables not in the model: score statistics and residual chi-square.

Cox Regression Data Considerations

Data. Your time variable should be quantitative and your status variable can be categorical or continuous. Independent variables (covariates) can be continuous or categorical; if categorical, they should be dummy- or indicator-coded (there is an option in the procedure to recode categorical variables automatically). Strata variables should be categorical, coded as integers or short strings.

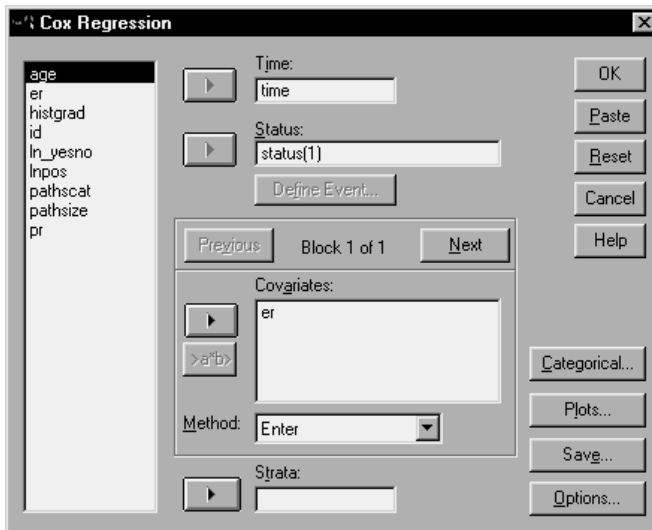
Assumptions. Observations should be independent, and the hazard ratio should be constant across time; that is, the proportionality of hazards from one case to another should not vary over time. The latter assumption is known as the **proportional hazards assumption**.

Related procedures. If the proportional hazards assumption does not hold (see above), you may need to use the Cox with Time-Dependent Covariates procedure. If you have no covariates, or if you have only one categorical covariate, you can use the Life Tables or Kaplan-Meier procedure to examine survival or hazard functions for your sample(s). If you have no censored data in your sample (that is, every case experienced the terminal event), you can use the Linear Regression procedure to model the relationship between predictors and time-to-event.

Obtaining a Cox Regression Analysis

- ▶ From the menus choose:
 - Analyze
 - Survival
 - Cox Regression...

Figure 11-1
Cox Regression dialog box

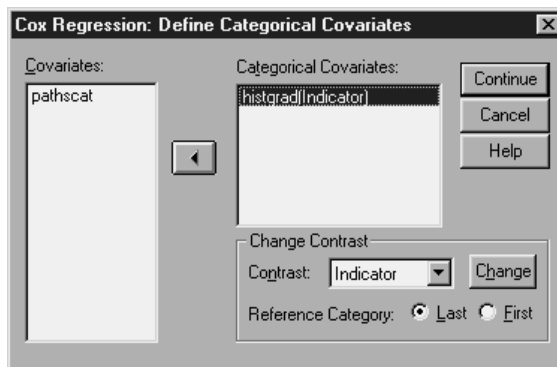


- ▶ Select a time variable. Cases whose time values are negative are not analyzed.
- ▶ Select a status variable, and then click Define Event.
- ▶ Select one or more covariates. To include interaction terms, select all of the variables involved in the interaction and then select >a*b>.

Optionally, you can compute separate models for different groups by defining a strata variable.

Cox Regression Define Categorical Variables

Figure 11-2
Cox Regression Define Categorical Covariates dialog box



You can specify details of how the Cox Regression procedure will handle categorical variables.

Covariates. Lists all of the covariates specified in the main dialog box, either by themselves or as part of an interaction, in any layer. If some of these are string variables or are categorical, you can use them only as categorical covariates.

Categorical Covariates. Lists variables identified as categorical. Each variable includes a notation in parentheses indicating the contrast coding to be used. String variables (denoted by the symbol < following their names) are already present in the Categorical Covariates list. Select any other categorical covariates from the Covariates list and move them into the Categorical Covariates list.

Change Contrast. Allows you to change the contrast method. Available contrast methods are:

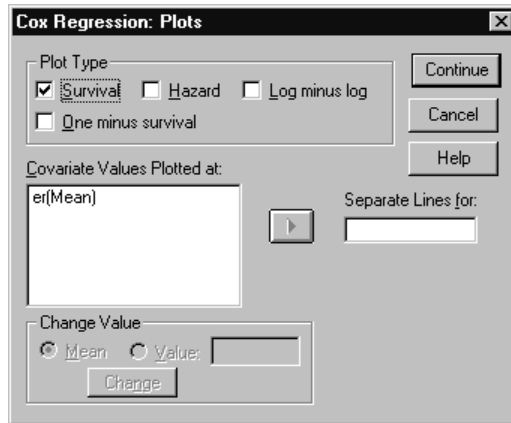
- **Indicator.** Contrasts indicate the presence or absence of category membership. The reference category is represented in the contrast matrix as a row of zeros.
- **Simple.** Each category of the predictor variable except the reference category is compared to the reference category.
- **Difference.** Each category of the predictor variable except the first category is compared to the average effect of previous categories. Also known as reverse Helmert contrasts.
- **Helmert.** Each category of the predictor variable except the last category is compared to the average effect of subsequent categories.
- **Repeated.** Each category of the predictor variable except the first category is compared to the category that precedes it.
- **Polynomial.** Orthogonal polynomial contrasts. Categories are assumed to be equally spaced. Polynomial contrasts are available for numeric variables only.
- **Deviation.** Each category of the predictor variable except the reference category is compared to the overall effect.

If you select Deviation, Simple, or Indicator, select either First or Last as the reference category. Note that the method is not actually changed until you click Change.

String covariates must be categorical covariates. To remove a string variable from the Categorical Covariates list, you must remove all terms containing the variable from the Covariates list in the main dialog box.

Cox Regression Plots

Figure 11-3
Cox Regression Plots dialog box



Plots can help you to evaluate your estimated model and interpret the results. You can plot the survival, hazard, log-minus-log, and one-minus-survival functions.

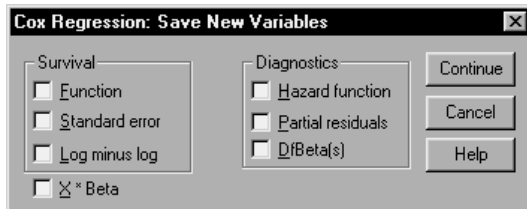
- **Survival Plot.** Displays the cumulative survival function on a linear scale.
- **Hazard Plot.** Displays the cumulative hazard function on a linear scale.
- **Log-Minus-Log.** The cumulative survival estimate after the $\ln(-\ln)$ transformation is applied to the estimate.
- **One Minus Survival.** Plots one-minus the survival function on a linear scale.

Because these functions depend on values of the covariates, you must use constant values for the covariates to plot the functions versus time. The default is to use the mean of each covariate as a constant value, but you can enter your own values for the plot using the Change Value control group.

You can plot a separate line for each value of a categorical covariate by moving that covariate into the Separate Lines For text box. This option is available only for categorical covariates, which are denoted by (Cat) after their names in the Covariate Values Plotted At list.

Cox Regression Save New Variables

Figure 11-4
Cox Regression Save New Variables dialog box



You can save various results of your analysis as new variables. These variables can then be used in subsequent analyses to test hypotheses or to check assumptions.

Survival. Allows you to save the survival function, standard error, and log-minus-log estimates as new variables.

Diagnostics. Allows you to save the hazard function, partial residuals, and DfBeta(s) for the regression as new variables.

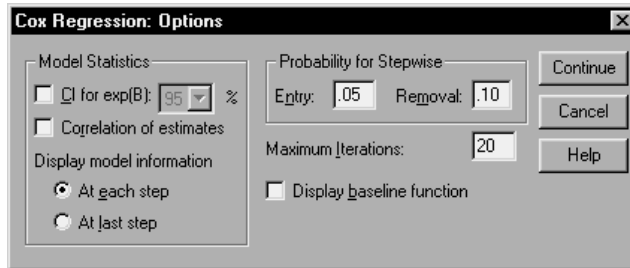
- **Hazard Function.** Saves the cumulative hazard function estimate (also called the Cox-Snell residual).
- **Partial Residuals.** You can plot partial residuals against survival time to test the proportional hazards assumption. One variable is saved for each covariate in the final model. Partial residuals are available only for models containing at least one covariate.
- **DfBetas.** Estimated change in a coefficient if a case is removed. One variable is saved for each covariate in the final model. DfBetas are only available for models containing at least one covariate.

If you are running Cox with a time-dependent covariate, DfBeta(s) are the only variables that you can save. You can also save the linear predictor variable X*Beta.

- **X*Beta.** Linear predictor score. The sum of the product of mean-centered covariate values and their corresponding parameter estimates for each case.

Cox Regression Options

Figure 11-5
Cox Regression Options dialog box



You can control various aspects of your analysis and output.

Model Statistics. You can obtain statistics for your model parameters, including confidence intervals for $\exp(B)$ and correlation of estimates. You can request these statistics either at each step or at the last step only.

Probability for Stepwise. If you have selected a stepwise method, you can specify the probability for either entry or removal from the model. A variable is entered if the significance level of its F -to-enter is less than the Entry value, and a variable is removed if the significance level is greater than the Removal value. The Entry value must be less than the Removal value.

Maximum Iterations. Allows you to specify the maximum iterations for the model, which controls how long the procedure will search for a solution.

Display baseline function. Allows you to display the baseline hazard function and cumulative survival at the mean of the covariates. This display is not available if you have specified time-dependent covariates.

Cox Regression Define Event for Status Variable

Enter the value or values indicating that the terminal event has occurred. You can enter a single value, a range of values, or a list of values. The Range of Values option is available only if your status variable is numeric.

COXREG Command Additional Features

The SPSS command language also allows you to:

- Obtain frequency tables that consider cases lost to follow-up as a separate category from censored cases.
- Select a reference category, other than first or last, for the deviation, simple, and indicator contrast methods.
- Specify unequal spacing of categories for the polynomial contrast method.
- Specify additional iteration criteria.
- Control the treatment of missing values.
- Specify the names for saved variables.
- Write output to an external SPSS system file.
- Hold data for each split-file group in an external scratch file during processing. This can help conserve memory resources when running analyses with large data sets. This is not available with time-dependent covariates.

See the *SPSS Command Syntax Reference* for complete syntax information.

Computing Time-Dependent Covariates

There are certain situations in which you would want to compute a Cox Regression model but the proportional hazards assumption does not hold. That is, hazard ratios change across time; the values of one (or more) of your covariates are different at different time points. In such cases, you need to use an extended Cox Regression model, which allows you to specify **time-dependent covariates**.

In order to analyze such a model, you must first define your time-dependent covariate. (Multiple time-dependent covariates can be specified using command syntax.) To facilitate this, a system variable representing time is available. This variable is called $T_$. You can use this variable to define time-dependent covariates in two general ways:

- If you want to test the proportional hazards assumption with respect to a particular covariate or estimate an extended Cox regression model that allows nonproportional hazards, you can do so by defining your time-dependent covariate as a function of the time variable $T_$ and the covariate in question. A common example would be the simple product of the time variable and the covariate, but more complex functions can be specified as well. Testing the significance of the coefficient of the time-dependent covariate will tell you whether the proportional hazards assumption is reasonable.
- Some variables may have different values at different time periods but aren't systematically related to time. In such cases, you need to define a **segmented time-dependent covariate**, which can be done using **logical expressions**. Logical expressions take the value 1 if true and 0 if false. Using a series of logical expressions, you can create your time-dependent covariate from a set of measurements. For example, if you have blood pressure measured once a week for the four weeks of your study (identified as $BP1$ to $BP4$), you can define your time-dependent covariate as $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq$

$2 \& T_ < 3) * BP3 + (T_ \geq 3 \& T_ < 4) * BP4$. Notice that exactly one of the terms in parentheses will be equal to 1 for any given case and the rest will all equal 0. In other words, this function means that if time is less than one week, use *BP1*; if it is more than one week but less than two weeks, use *BP2*, and so on.

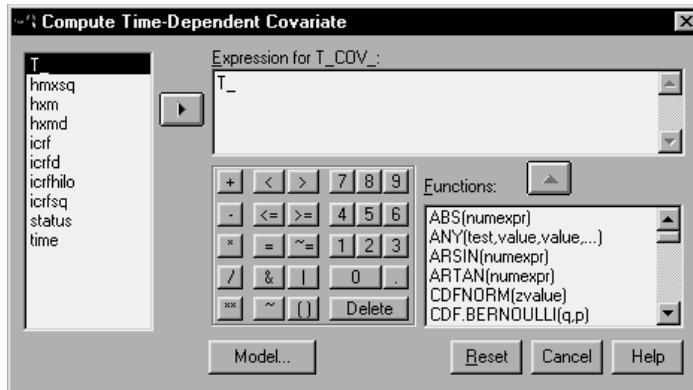
For segmented time-dependent covariates, cases that are missing any values are removed from the analysis. Therefore, you must be sure that all cases have values for all measured time points on the covariate, even for time points after the case is removed from the risk set (due to event or censoring). These values are not used in the analysis, but they must be valid SPSS values to prevent the cases from being dropped. For example, with the definition given above, a case censored at the second week must still have values for *BP3* and *BP4* (they can be 0 or any other number, since they are not used in the analysis).

In the Compute Time-Dependent Covariate dialog box, you can use the function-building controls to build the expression for the time-dependent covariate, or you can enter it directly in the Expression for T_COV_ text area. Note that string constants must be enclosed in quotation marks or apostrophes, and numeric constants must be typed in American format, with the dot as the decimal delimiter. The resulting variable is called *T_COV_* and should be included as a covariate in your Cox Regression model.

Computing a Time-Dependent Covariate

- ▶ From the menus choose:
 - Analyze
 - Survival
 - Cox w/ Time-Dep Cov...

Figure 12-1
 Compute Time-Dependent Covariate dialog box



- ▶ Enter an expression for the time-dependent covariate.
- ▶ Click Model to proceed with your Cox Regression.

Note: Be sure to include the new variable *T_COV_* as a covariate in your Cox Regression model.

For more information, see “Cox Regression Analysis” in Chapter 11 on page 105.

Cox Regression with Time-Dependent Covariates Additional Features

The SPSS command language also allows you to specify multiple time-dependent covariates.

Other command syntax features are available for Cox Regression with or without time-dependent covariates. See the *SPSS Command Syntax Reference* for complete syntax information.

Categorical Variable Coding Schemes

In many SPSS procedures, you can request automatic replacement of a categorical independent variable with a set of contrast variables, which will then be entered or removed from an equation as a block. You can specify how the set of contrast variables is to be coded, usually on the CONTRAST subcommand. This appendix explains and illustrates how different contrast types requested on CONTRAST actually work.

Deviation

Deviation from the grand mean. In matrix terms, these contrasts have the form:

$$\begin{array}{l}
 \text{mean} \quad (1/k \quad 1/k \quad \dots \quad 1/k \quad 1/k) \\
 \text{df(1)} \quad (1-1/k \quad -1/k \quad \dots \quad -1/k \quad -1/k) \\
 \text{df(2)} \quad (-1/k \quad 1-1/k \quad \dots \quad -1/k \quad -1/k) \\
 \quad \cdot \quad \quad \quad \cdot \\
 \quad \cdot \quad \quad \quad \cdot \\
 \text{df}(k-1) \quad (-1/k \quad -1/k \quad \dots \quad 1-1/k \quad -1/k)
 \end{array}$$

where k is the number of categories for the independent variable and the last category is omitted by default. For example, the deviation contrasts for an independent variable with three categories are as follows:

$$\begin{array}{l}
 (1/3 \quad 1/3 \quad 1/3) \\
 (2/3 \quad -1/3 \quad -1/3) \\
 (-1/3 \quad 2/3 \quad -1/3)
 \end{array}$$

To omit a category other than the last, specify the number of the omitted category in parentheses after the DEVIATION keyword. For example, the following subcommand obtains the deviations for the first and third categories and omits the second:

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

Suppose that *factor* has three categories. The resulting contrast matrix will be

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$$

Simple

Simple contrasts. Compares each level of a factor to the last. The general matrix form is

$$\begin{array}{l} \text{mean} \\ \text{df}(1) \\ \text{df}(2) \\ \cdot \\ \cdot \\ \text{df}(k-1) \end{array} \begin{pmatrix} 1/k & 1/k & \dots & 1/k & 1/k \\ 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

where k is the number of categories for the independent variable. For example, the simple contrasts for an independent variable with four categories are as follows:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

To use another category instead of the last as a reference category, specify in parentheses after the SIMPLE keyword the sequence number of the reference category, which is not necessarily the value associated with that category. For example, the

following CONTRAST subcommand obtains a contrast matrix that omits the second category:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Suppose that *factor* has four categories. The resulting contrast matrix will be

```
( 1/4    1/4    1/4    1/4 )
( 1     -1     0     0 )
( 0     -1     1     0 )
( 0     -1     0     1 )
```

Helmert

Helmert contrasts. Compares categories of an independent variable with the mean of the subsequent categories. The general matrix form is

```
mean      ( 1/k    1/k    ...    1/k    1/k )
df(1)     ( 1   -1/(k-1)  ...  -1/(k-1)  -1/(k-1) )
df(2)     ( 0     1     ...  -1/(k-2)  -1/(k-2) )
.         .
.         .
df(k-2)   ( 0     0     1    -1/2    -1/2 )
df(k-1)   ( 0     0     ...    1     -1 )
```

where k is the number of categories of the independent variable. For example, an independent variable with four categories has a Helmert contrast matrix of the following form:

```
( 1/4    1/4    1/4    1/4 )
( 1     -1/3   -1/3   -1/3 )
( 0     1     -1/2   -1/2 )
( 0     0     1     -1 )
```

Difference

Difference or reverse Helmert contrasts. Compares categories of an independent variable with the mean of the previous categories of the variable. The general matrix form is

mean	(1/k	1/k	1/k	...	1/k)
df(1)	(-1	1	0	...	0)
df(2)	(-1/2	-1/2	1	...	0)
.
.
df(k-1)	(-1/(k-1)	-1/(k-1)	-1/(k-1)	...	1)

where k is the number of categories for the independent variable. For example, the difference contrasts for an independent variable with four categories are as follows:

(1/4	1/4	1/4	1/4)
(-1	1	0	0)
(-1/2	-1/2	1	0)
(-1/3	-1/3	-1/3	1)

Polynomial

Orthogonal polynomial contrasts. The first degree of freedom contains the linear effect across all categories; the second degree of freedom, the quadratic effect; the third degree of freedom, the cubic; and so on, for the higher-order effects.

You can specify the spacing between levels of the treatment measured by the given categorical variable. Equal spacing, which is the default if you omit the metric, can be specified as consecutive integers from 1 to k , where k is the number of categories. If the variable *drug* has three categories, the subcommand

```
/CONTRAST(DRUG)=POLYNOMIAL
```

is the same as

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Equal spacing is not always necessary, however. For example, suppose that *drug* represents different dosages of a drug given to three groups. If the dosage administered to the second group is twice that given to the first group and the dosage administered to the third group is three times that given to the first group, the treatment categories are equally spaced, and an appropriate metric for this situation consists of consecutive integers:

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

If, however, the dosage administered to the second group is four times that given to the first group, and the dosage administered to the third group is seven times that given to the first group, an appropriate metric is

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

In either case, the result of the contrast specification is that the first degree of freedom for *drug* contains the linear effect of the dosage levels and the second degree of freedom contains the quadratic effect.

Polynomial contrasts are especially useful in tests of trends and for investigating the nature of response surfaces. You can also use polynomial contrasts to perform nonlinear curve fitting, such as curvilinear regression.

Repeated

Compares adjacent levels of an independent variable. The general matrix form is

mean	(1/k	1/k	1/k	...	1/k	1/k)
df(1)	(1	-1	0	...	0	0)
df(2)	(0	1	-1	...	0	0)
.			.					
.			.					
df(k-1)	(0	0	0	...	1	-1)

where *k* is the number of categories for the independent variable. For example, the repeated contrasts for an independent variable with four categories are as follows:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

These contrasts are useful in profile analysis and wherever difference scores are needed.

Special

A user-defined contrast. Allows entry of special contrasts in the form of square matrices with as many rows and columns as there are categories of the given independent variable. For MANOVA and LOGLINEAR, the first row entered is always the mean, or constant, effect and represents the set of weights indicating how to average other independent variables, if any, over the given variable. Generally, this contrast is a vector of ones.

The remaining rows of the matrix contain the special contrasts indicating the desired comparisons between categories of the variable. Usually, orthogonal contrasts are the most useful. Orthogonal contrasts are statistically independent and are nonredundant. Contrasts are orthogonal if:

- For each row, contrast coefficients sum to 0.
- The products of corresponding coefficients for all pairs of disjoint rows also sum to 0.

For example, suppose that treatment has four levels and that you want to compare the various levels of treatment with each other. An appropriate special contrast is

$(1 \quad 1 \quad 1 \quad 1)$	weights for mean calculation
$(3 \quad -1 \quad -1 \quad -1)$	compare 1 st with 2 nd through 4 th
$(0 \quad 2 \quad -1 \quad -1)$	compare 2 nd with 3 rd and 4 th
$(0 \quad 0 \quad 1 \quad -1)$	compare 3 rd with 4 th

which you specify by means of the following CONTRAST subcommand for MANOVA, LOGISTICREGRESSION, and COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1  1  1  1
                               3 -1 -1 -1
                               0  2 -1 -1
                               0  0  1 -1 )
```

For LOGLINEAR, you need to specify:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1  1  1  1
                                     3 -1 -1 -1
                                     0  2 -1 -1
                                     0  0  1 -1 )
```

Each row except the means row sums to 0. Products of each pair of disjoint rows sum to 0 as well:

$$\text{Rows 2 and 3:} \quad (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$$

$$\text{Rows 2 and 4:} \quad (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$$

$$\text{Rows 3 and 4:} \quad (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0$$

The special contrasts need not be orthogonal. However, they must not be linear combinations of each other. If they are, the procedure reports the linear dependency and ceases processing. Helmert, difference, and polynomial contrasts are all orthogonal contrasts.

Indicator

Indicator variable coding. Also known as dummy coding, this is not available in LOGLINEAR or MANOVA. The number of new variables coded is $k-1$. Cases in the reference category are coded 0 for all $k-1$ variables. A case in the i^{th} category is coded 0 for all indicator variables except the i^{th} , which is coded 1.

Covariance Structures

This section provides additional information on covariance structures.

Ante-Dependence: First-Order. This covariance structure has heterogenous variances and heterogenous correlations between adjacent elements. The correlation between two non-adjacent elements is the product of the correlations between the elements that lie between the elements of interest.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\ \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$$

AR(1). This is a first-order autoregressive structure with homogenous variances. The correlation between any two elements is equal to rho for adjacent elements, rho² for elements that are separated by a third, and so on. Rho is constrained so that -1 < rho < 1.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

AR(1): Heterogenous. This is a first-order autoregressive structure with heterogenous variances. The correlation between any two elements is equal to rho for adjacent elements, rho² for two elements separated by a third, and so on. Rho is constrained to lie between -1 and 1.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

ARMA(1,1). This is a first-order autoregressive moving average structure. It has homogenous variances. The correlation between two elements is equal to $\phi\rho$ for adjacent elements, $\phi^2\rho^2$ for elements separated by a third, and so on. ρ and ϕ are the autoregressive and moving average parameters, respectively, and their values are constrained to lie between -1 and 1 , inclusive.

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi^2\rho^2 & \phi^3\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi^2\rho^2 \\ \phi^2\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi^3\rho^3 & \phi^2\rho^2 & \phi\rho & 1 \end{bmatrix}$$

Compound Symmetry. This structure has constant variance and constant covariance.

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

Compound Symmetry: Correlation Metric. This covariance structure has homogenous variances and homogenous correlations between elements.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Compound Symmetry: Heterogenous. This covariance structure has heterogenous variances and constant correlation between elements.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho & \sigma_4\sigma_1\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

Diagonal. This covariance structure has heterogenous variances and zero correlation between elements.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Factor Analytic: First-Order. This covariance structure has heterogenous variances that are composed of a term that is heterogenous across elements and a term that is homogenous across elements. The covariance between any two elements is the square root of the product of their heterogenous variance terms.

$$\begin{bmatrix} \lambda_1^2 + d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d \end{bmatrix}$$

Factor Analytic: First-Order, Heterogenous. This covariance structure has heterogenous variances that are composed of two terms that are heterogenous across elements. The covariance between any two elements is the square root of the product of the first of their heterogenous variance terms.

$$\begin{bmatrix} \lambda_1^2 + d_1 & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d_2 & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d_3 & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d_4 \end{bmatrix}$$

Huynh-Feldt. This is a “circular” matrix in which the covariance between any two elements is equal to the average of their variances minus a constant. Neither the variances nor the covariances are constant.

$$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$$

Scaled Identity. This structure has constant variance. There is assumed to be no correlation between any elements.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Toeplitz. This covariance structure has homogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Toeplitz: Heterogenous. This covariance structure has heterogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_1 & \sigma_3 \sigma_1 \rho_2 & \sigma_4 \sigma_1 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_3 \sigma_2 \rho_1 & \sigma_4 \sigma_2 \rho_2 \\ \sigma_3 \sigma_1 \rho_2 & \sigma_3 \sigma_2 \rho_1 & \sigma_3^2 & \sigma_4 \sigma_3 \rho_1 \\ \sigma_4 \sigma_1 \rho_3 & \sigma_4 \sigma_2 \rho_2 & \sigma_4 \sigma_3 \rho_1 & \sigma_4^2 \end{bmatrix}$$

Unstructured. This is a completely general covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Unstructured: Correlation Metric. This covariance structure has heterogenous variances and heterogenous correlations.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$$

Variance Components. This structure assigns a scaled identity (ID) structure to each of the specified random effects.

- analysis of covariance, 1
 - in GLM Multivariate, 1
- analysis of variance, 39
 - in Variance Components, 39
- ANOVA, 1, 17, 20
 - in GLM Multivariate, 1
 - in GLM Repeated Measures, 17, 20
- backward elimination, 59
 - in Model Selection Loglinear Analysis, 59
- Bartlett's test of sphericity, 13
 - in GLM Multivariate, 13
- Box's M test, 13
 - in GLM Multivariate, 13
- Breslow test, 100
 - in Kaplan-Meier, 100
- categorical covariates, 107
- censored cases, 91, 97, 105
 - in Cox Regression, 105
 - in Kaplan-Meier, 97
 - in Life Tables, 91
- confidence intervals, 13, 13, 32, 32, 55, 69, 78
 - in General Loglinear Analysis, 69
 - in GLM Multivariate, 13, 13
 - in GLM Repeated Measures, 32, 32
 - in Linear Mixed Models, 55
 - in Logit Loglinear Analysis, 78
- contingency tables, 59, 65
 - in General Loglinear Analysis, 65
- contrasts, 65, 73, 73, 107
 - in Cox Regression, 107
 - in General Loglinear Analysis, 65
 - in Logit Loglinear Analysis, 73
- Cook's distance, 30
 - in GLM Repeated Measures, 30
- correlation matrix, 55, 85
 - in Linear Mixed Models, 55
 - in Ordinal Regression, 85
- covariance matrix, 55, 85
 - in Linear Mixed Models, 55
 - in Ordinal Regression, 85
- covariance parameters test, 55
 - in Linear Mixed Models, 55
- covariance structures, 125
 - in Linear Mixed Models, 125
- covariates
 - in Cox Regression, 107
- Cox and Snell R-square, 85
 - in Ordinal Regression, 85
- Cox Regression, 105, 113
 - assumptions, 105
 - baseline functions, 111
 - categorical covariates, 107
 - command additional features, 112
 - contrasts, 107
 - covariates, 105
 - data, 105
 - define event, 111
 - DfBeta(s), 110
 - example, 105
 - hazard function, 110
 - iterations, 111
 - partial residuals, 110
 - plots, 109
 - related procedures, 105
 - saving new variables, 110
 - statistics, 105, 111
 - stepwise entry and removal, 111
 - string covariates, 107
 - survival function, 110
 - survival status variable, 111
 - time-dependent covariates, 113, 115
- cross-products, 13
 - hypothesis and error matrices, 13
- crossstabulation
 - in Model Selection Loglinear Analysis, 59
- cumulative frequencies, 85
 - in Ordinal Regression, 85
- custom models, 23, 38, 61
 - in GLM Repeated Measures, 23
 - in Model Selection Loglinear Analysis, 61
 - in Variance Components, 38
- deleted residuals, 30
 - in GLM Repeated Measures, 30

- descriptive statistics, 13, 32, 55
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - in Linear Mixed Models, 55
- effect-size estimates, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- estimated marginal means, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - in Linear Mixed Models, 56
- eta-squared, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- expected frequencies, 85
 - in Ordinal Regression, 85
- factor-level information, 55
 - in Linear Mixed Models, 55
- Fisher scoring, 53
 - in Linear Mixed Models, 53
- fixed effects, 48
 - in Linear Mixed Models, 48
- fixed predicted values
 - in Linear Mixed Models, 57
- frequencies, 63
 - in Model Selection Loglinear Analysis, 63
- full factorial models, 23, 38
 - in GLM Repeated Measures, 23
 - in Variance Components, 38
- Gehan test, 95
 - in Life Tables, 95
- generalized log-odds ratio, 65
 - in General Loglinear Analysis, 65
- General Loglinear Analysis, 65, 66, 66, 66, 68, 69, 70, 71
 - assumptions, 66
 - cell covariates, 65
 - cell structures, 65
 - confidence intervals, 69
 - contrasts, 65
 - criteria, 69
 - display options, 69
 - distribution of cell counts, 65
 - factors, 65
 - model specification, 68
 - multinomial distribution, 66
 - plots, 69
 - Poisson distribution, 66
 - residuals, 70
 - saving predicted values, 70
 - saving variables, 70
- generating class, 61
 - in Model Selection Loglinear Analysis, 61
- GLM Multivariate, 1, 2, 3, 13, 14
 - assumptions, 2
 - covariates, 1
 - dependent variable, 1
 - diagnostics, 13
 - display, 13
 - estimated marginal means, 13
 - factors, 1
 - options, 13
- GLM Repeated Measures, 17, 18, 20, 21, 22, 23, 30, 32, 32, 33
 - assumptions, 18
 - diagnostics, 32
 - display, 32
 - estimated marginal means, 32
 - model, 23
 - options, 32
 - saving variables, 30
- GLOR, 65
 - in General Loglinear Analysis, 65
- goodness of fit, 85
 - in Ordinal Regression, 85
- hazard rate, 91
 - in Life Tables, 91
- hierarchical decomposition
 - in GLM, 40
- hierarchical loglinear models, 59
- homogeneity-of-variance tests, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- interaction terms, 48
 - in Linear Mixed Models, 48
- iteration history, 53, 85
 - in Linear Mixed Models, 53
 - in Ordinal Regression, 85

- iterations, 63
 - in Model Selection Loglinear Analysis, 63
- Kaplan-Meier, 97
 - assumptions, 97
 - command additional features, 103
 - comparing factor levels, 100
 - data, 97
 - defining events, 100
 - example, 97
 - linear trend for factor levels, 100
 - mean and median survival time, 102
 - plots, 102
 - quartiles, 102
 - related procedures, 97
 - saving new variables, 101
 - statistics, 97, 102
 - survival status variables, 100
 - survival tables, 102
- Levene test, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- leverage values, 30
 - in GLM Repeated Measures, 30
- Life Tables, 91
 - assumptions, 92
 - censored cases, 92
 - command additional features, 96
 - comparing factor levels, 95
 - data considerations, 92
 - example, 91
 - factor variables, 94
 - hazard rate, 91
 - plots, 95
 - related procedures, 92
 - statistics, 91
 - suppressing table display, 95
 - survival function, 91
 - survival status variables, 94
 - Wilcoxon (Gehan) test, 95
- likelihood-ratio chi-square statistic, 85
 - in Ordinal Regression, 85
- Linear Mixed Models, 43, 44, 46, 47, 48, 48, 49, 51, 53, 55, 56, 57, 58, 125
 - assumptions, 44
 - build terms, 48, 49
 - covariance structure, 125
 - estimated marginal means, 56
 - estimation criteria, 53
 - fixed effects, 48
 - interaction terms, 48
 - model, 55
 - random effects, 51
 - save variables, 57
- link
 - in Ordinal Regression, 83
- location model, 86
 - in Ordinal Regression, 86
- Logit Loglinear Analysis, 73, 73, 74, 74, 76, 78, 79
 - assumptions, 74
 - cell covariates, 73
 - cell structures, 73
 - confidence intervals, 78
 - contrasts, 73
 - criteria, 78
 - display options, 78
 - distribution of cell counts, 73
 - factors, 73
 - model specification, 76
 - plots, 78
 - predicted values, 79
 - residuals, 79
 - saving variables, 79
- log-likelihood convergence, 53
 - in Linear Mixed Models, 53
- loglinear analysis, 59, 65, 73
 - General Loglinear Analysis, 65
 - Logit Loglinear Analysis, 73
- log rank test, 100
 - in Kaplan-Meier, 100
- Mauchly's test of sphericity, 18, 32
 - in GLM Repeated Measures, 18, 32
- maximum likelihood estimation, 39
 - in Variance Components, 39
- McFadden R-square, 85
 - in Ordinal Regression, 85
- MINQUE, 39
 - in Variance Components, 39
- mixed models, 18, 43
 - linear, 43

- Model Selection Loglinear Analysis, 59, 59, 60, 61, 61, 63, 63
 - assumptions, 59
 - defining factor ranges, 61
 - models, 61
 - options, 63
- multinomial distribution, 66, 74
 - in General Loglinear Analysis, 66
 - in Logit Loglinear Analysis, 74
- multinomial logit models, 73
- multivariate ANOVA, 1
- multivariate GLM, 1
- Nagelkerke R-square, 85
 - in Ordinal Regression, 85
- nested terms, 49
 - in Linear Mixed Models, 49
- Newton-Raphson method, 65, 73
 - in General Loglinear Analysis, 65
 - in Logit Loglinear Analysis, 73
- normal probability plots, 63
 - in Model Selection Loglinear Analysis, 63
- observed frequencies, 85
 - in Ordinal Regression, 85
- observed means, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- odds ratio, 65
 - in General Loglinear Analysis, 65
- Ordinal Regression, 81, 82, 82, 83, 85, 86, 88, 89
 - assumptions, 82
 - link, 83
 - options, 83
 - statistics, 81
- parameter convergence, 53
 - in Linear Mixed Models, 53
- parameter covariance matrix, 55
 - in Linear Mixed Models, 55
- parameter estimates, 13, 32, 55, 63, 65, 73, 85
 - in General Loglinear Analysis, 65
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - in Linear Mixed Models, 55
 - in Logit Loglinear Analysis, 73
 - in Model Selection Loglinear Analysis, 63
 - in Ordinal Regression, 85
- Pearson chi-square, 85
 - in Ordinal Regression, 85
- Pearson residuals, 85
 - in Ordinal Regression, 85
- plots, 69, 78
 - in General Loglinear Analysis, 69
 - in Logit Loglinear Analysis, 78
- PLUM
 - in Ordinal Regression, 81
- Poisson distribution, 66
 - in General Loglinear Analysis, 66
- power estimates, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- predicted values, 70, 79
 - in General Loglinear Analysis, 70
 - in Linear Mixed Models, 57
 - in Logit Loglinear Analysis, 79
- proportional hazards model, 105
 - in Cox Regression, 105
- random-effect covariance matrix, 55
 - in Linear Mixed Models, 55
- random-effect priors, 39
 - in Variance Components, 39
- random effects, 51
 - in Linear Mixed Models, 51
- repeated measures variables, 44
 - in Linear Mixed Models, 44
- residual covariance matrix, 55
 - in Linear Mixed Models, 55
- residual plots, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- residuals, 63, 70, 79
 - in General Loglinear Analysis, 70
 - in Linear Mixed Models, 57
 - in Logit Loglinear Analysis, 79
 - in Model Selection Loglinear Analysis, 63
- residual SSCP, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
- restricted maximum likelihood estimation, 39
 - in Variance Components, 39

-
- saturated models, 61
 - in Model Selection Loglinear Analysis, 61
 - scale model, 88
 - in Ordinal Regression, 88
 - scoring, 53
 - in Linear Mixed Models, 53
 - segmented time-dependent covariates, 113
 - in Cox Regression, 113
 - singularity tolerance, 53
 - in Linear Mixed Models, 53
 - split-plot analysis, 18
 - spread-versus-level plots, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - SSCP, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - standard deviation, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - standard error, 13, 30, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 30, 32
 - standardized residuals, 30
 - in GLM Repeated Measures, 30
 - step-halving, 53
 - in Linear Mixed Models, 53
 - string covariates, 107
 - in Cox Regression, 107
 - subjects variables, 44
 - in Linear Mixed Models, 44
 - sum of squares, 40, 49
 - in Linear Mixed Models, 49
 - in Variance Components, 40
 - sums of squares, 13
 - hypothesis and error matrices, 13
 - survival analysis, 91, 97, 105, 113
 - in Cox Regression, 105
 - in Kaplan-Meier, 97
 - in Life Tables, 91
 - Time-Dependent Cox Regression, 113
 - survival function, 91
 - in Life Tables, 91
 - Tarone-Ware test, 100
 - in Kaplan-Meier, 100
 - test of parallel lines, 85
 - in Ordinal Regression, 85
 - t tests, 13, 32
 - in GLM Multivariate, 13
 - in GLM Repeated Measures, 32
 - unstandardized residuals, 30
 - in GLM Repeated Measures, 30
 - Variance Components, 35, 36, 36, 38, 39, 41, 42
 - assumptions, 36
 - model, 38
 - options, 39
 - saving results, 41
 - Wald statistic, 65, 73
 - in General Loglinear Analysis, 65
 - in Logit Loglinear Analysis, 73
 - weighted predicted values, 30
 - in GLM Repeated Measures, 30
 - Wilcoxon test, 95
 - in Life Tables, 95