Topics in Survey Methodology and Survey Analysis, fall 2013

Part 3: Risto Lehtonen
Tuesday 24 Sept. at 16-20
Thursday 26 Sept. at 14-19

Collection of supplemental materials

Survey analysis software: SAS, SPSS, Mplus

TABLE: GENERALIZED LINEAR (MIXED) MODELS IN SAS

SAS model-based procedures

SAS SURVEYMEANS, subgroup means for OHC survey

SAS SURVEYLOGISTIC, Logistic ANCOVA

PC Session, SAS1 code: Descriptives and design-based tests

PC Session, SAS2 code: Logistic regression for cluster correlated data

SPSS CSPLAN file for OHC survey

SPSS Complex Samples, Subgroup means for OHC survey

SPSS Complex Samples, Test of independence for OHC survey

SPSS Complex Samples, Logistic ANCOVA

PC Session, Mplus Logistic Regression for OHC Survey data: COMPLEX type analysis for cluster correlated data

SAS - Sample Survey Design and Analysis: Overview

Researchers often use sample survey methodology to obtain information about a large aggregate or population by selecting and measuring a sample from that population. Due to the variability of characteristics among items in the population, researchers apply scientific sample designs in the sample selection process to reduce the risk of a distorted view of the population, and they make inferences about the population based on the information from the sample survey data. In order to make statistically valid inferences for the population, they must incorporate the sample design in the data analysis.

Traditional SAS® procedures, such as the MEANS procedure and the GLM procedure, compute statistics under the assumption that the sample is drawn from an infinite population by simple random sampling. These procedures generally do not correctly estimate the variance of an estimator if they are applied to a sample drawn by a complex sample design. SAS users have requested procedures that analyze data from complex sample surveys. In response to this request, SAS/STAT® software now provides the SURVEYFREQ, SURVEYLOGISTIC, SURVEYMEANS, SURVEYPHREG, SURVEYREG, and SURVEYSELECT procedures.

To select probability-based random samples from a study population, you can use the SURVEYSELECT procedure, which provides a variety of methods for probability sampling. To analyze sample survey data, you can use the SURVEYFREQ, SURVEYLOGISTIC, SURVEYMEANS, SURVEYPHREG, and SURVEYREG procedures, which incorporate the sample design into the analyses. These procedures can be used for multistage designs or for single-stage designs, with or without stratification, and with or without unequal weighting.

Procedure	Features
PROC SURVEYFREQ	estimates of population means and totals, estimates of population proportions, standard errors, confidence limits, hypothesis tests (t tests), domain analysis, ratio estimates
PROC SURVEYLOGISTIC	cumulative logit regression model fitting, logit, complementary log-log and probit link functions, generalized logit regression model fitting, estimates of regression coefficients, estimates of covariance matrices, hypothesis tests, model diagnostics, estimates of odds ratios, confidence limits, estimable functions, estimates and standard errors for contrasts, domain analysis
PROC SURVEYMEANS	estimates of population means and totals, estimates of population proportions, standard errors, confidence limits, hypothesis tests (t tests), domain analysis, ratio estimates
PROC SURVEYPHREG	regression analysis based on the Cox proportional hazards model, hazard ratio estimates, predicted values and their standard errors, martingale, Schoenfeld, score, and deviance residuals, significance tests, confidence limits, estimable functions, domain analysis
PROC SURVEYREG	linear regression model fitting, estimates of regression coefficients, estimates of covariance matrices, significance tests, confidence limits, estimable functions, estimates and standard errors for contrasts, domain analysis,
PROC SURVEYSELECT	simple random sampling, unrestricted random sampling (with replacement), systematic sampling, sequential sampling, selection probability proportional to size (PPS) with and without replacement, PPS systematic sampling, PPS for two units per stratum, sequential PPS with minimum replacement

Mplus

TUESDAY
SEPTEMBER 25, 2012

SEARCH
Go

MODELING WITH COMPLEX SURVEY DATA

There are two approaches to the analysis of complex survey data in Mplus. One approach is to compute standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations due to cluster sampling, and/or unequal probability of selection. Subpopulation analysis, replicate weights, and finite population correction are also available. With sampling weights, parameters are estimated by maximizing a weighted loglikelihood function. Standard error computations use a sandwich estimator. For this approach, observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or combinations of these variable types.

A second approach is to specify a model for each level of the multilevel data thereby modeling the non-independence of observations due to cluster sampling. This is commonly referred to as multilevel modeling. The use of sampling weights in the estimation of parameters, standard errors, and the chi-square test of model fit is allowed. Both individual-level and cluster-level weights can be used. With sampling weights, parameters are estimated by maximizing a weighted loglikelihood function. Standard error computations use a sandwich estimator. For this approach, observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or combinations of these variable types.

The multilevel extension of the full modeling framework allows random intercepts and random slopes that vary across clusters in hierarchical data. These random effects can be specified for any of the relationships of the full Mplus model for both independent and dependent variables and both observed and latent variables. Random effects representing across-cluster variation in intercepts and slopes or individual differences in growth can be combined with factors measured by multiple indicators on both the individual and cluster levels. In line with SEM, regressions among random effects, among factors, and between random effects and factors are allowed.

The two approaches described above can be combined. In addition to specifying a model for each level of the multilevel data thereby modeling the non-independence of observations due to cluster sampling, standard errors and a chi-square test of model fit are computed taking into account stratification, non-independence of observations due to cluster sampling, and/or unequal probability of selection. When there is clustering due to both primary and secondary sampling stages, the standard errors and chi-square test of model fit are computed taking into account the clustering due to the primary sampling stage and clustering due to the secondary sampling stage is modeled.

Most of the special features listed above are available for modeling of complex survey data.

Modeling with Both Continuous and Categorical Latent Variables
Back

Modeling with Missing Data
Next

SPSS Complex Samples

Work efficiently and easily with complex sample survey results

Only IBM SPSS Complex Samples makes understanding and working with your complex sample survey results easy. Through the intuitive interface, you can analyze data and interpret results. Choose from one of several wizards to make it easier to create plans, analyze data and interpret results.

When you're finished, you can publish public-use datasets and include your sampling and analysis plans. These plans act as a template and allow you to save all the decisions made when creating the plan – define it once and you're done. This saves time and improves accuracy for yourself and others who may want to plug your plans into the data to replicate results or pick up where you left off.

Use the following types of sample design information with IBM SPSS Complex Samples:

- **Stratified sampling** - Increase the precision of your sample or ensure a representative sample from key groups by choosing to sample within subgroups of the survey population.
- **Clustered sampling** - Select clusters, which are groups of sampling units, for your survey. Clustering often helps makes surveys more cost-effective.
- **Multistage sampling** - Select an initial or first-stage sample based on groups of elements in your population; then create a second-stage sample by drawing a sub-sample from each selected unit in the first-stage sample. By repeating this option, you can select a higher-stage sample.

Everything You Need for Planning

To help you through the planning stage in the analytical process, IBM SPSS Complex Samples provides you with specialized tools and procedures for working with sample survey data:

- **IBM SPSS Complex Samples Plan** (CSPLAN) – Use this procedure to specify the sampling frame to create a complex sample design or analysis specification used by companion procedures in IBM SPSS Complex Samples.
- **Sampling Plan Wizard** – If you are creating your own samples, use the Sampling Plan Wizard to define the scheme and draw the sample.
- **Analysis Preparation Wizard** – If you're using public-use datasets that already have samples, use the Analysis Plan Wizard to specify how the samples were defined and how standard errors should be estimated.
- **Plan files** – Once you have created plan files, you can save them and treat them as templates. This allows you to save all the decisions you made when creating the plan. This saves time and improves accuracy for yourself and others who may want to plug your plans into the data to replicate results or pick up where you left off.

Everything You Need for Data Management

IBM SPSS Complex Samples provides what you need for the data management stage when working with sample survey data. And it easily plugs into other IBM SPSS Statistics modules so you can seamlessly work in the IBM SPSS Statistics environment.

IBM SPSS Complex Samples Selection (CSSELECT) procedure – Enables you to select complex, probability-based samples from a population while mitigating the risk in doing so (e.g. over- or under-representing a subgroup). CSSELECT chooses units according to a sample design created through the CSPLAN procedure.

With this procedure, you can:

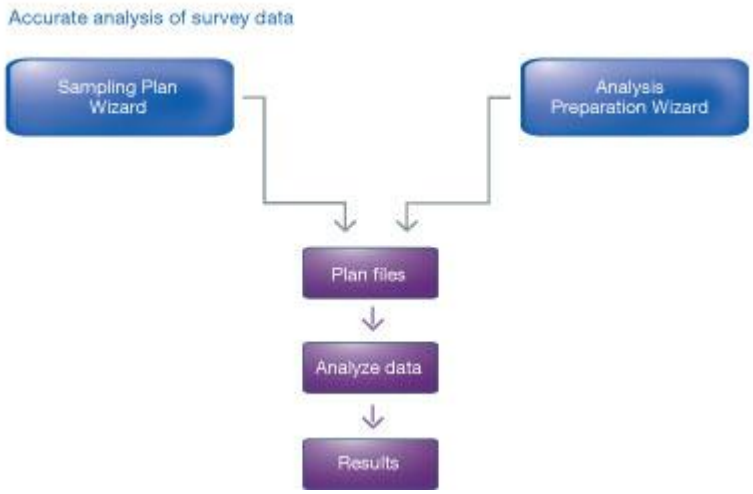
- Control the scope of execution and specify a seed value with the CRITERIA subcommand
- Control whether or not user-missing values of classification (stratification and clustering) variables are treated as valid variables with the CLASSMISSING subcommand
- Specify general options concerning input and output files with the DATA subcommand
- Write sampled units to an external file using an option to keep/drop specified variables
- Automatically save first-stage joint inclusion probabilities to an external file when the plan specifies a probability proportionate to size (PPS) without replacement (WR) sampling method
- Opt to generate text files containing a rule that describes characteristics of selected units

Everything You Need for Data Analysis

Performing data analysis in IBM SPSS Complex Samples helps you to achieve more statistically valid inferences for populations measured in your complex sample data. IBM SPSS Complex Samples provides you with better results because, unlike most conventional statistical software, it incorporates the sample design into survey analysis.

IBM SPSS Complex Samples features five procedures to analyze data from sample survey data:

- **IBM SPSS Complex Samples Descriptives** (CSDESCRIPTIVES) – Estimates means, sums and ratios, and computes standard errors, design effects, confidence intervals hypothesis tests for samples drawn by complex methods.
- **Complex Sample Tabulate** (CSTABULATE) – Displays one-way frequency tables or two-way crosstabulations and associated standard errors, design effects, confidence intervals and hypothesis tests for samples drawn by complex sampling methods.
- **IBM SPSS Complex Samples General Linear Models** (CSGLM) – Enables you to build linear regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) models for samples drawn by complex sampling methods.
- **IBM SPSS Complex Samples Logistic Regression** (CSLOGISTIC) – Performs binary logistic regression analysis, as well as multiple logistic regression (MLR) analysis, for samples drawn by complex sampling methods.
- **IBM SPSS Complex Samples Cox Regression** (CSCOXREG) – Applies Cox proportional hazards regression to analysis of survival times; that is, the length of time before the occurrence of an event for samples drawn by complex sampling methods.



Accurate analysis of survey data is easy in IBM SPSS Complex Samples. Start with one of the wizards (which one depends on your data source) and then use the interactive interface to create plans, analyze data and interpret results.

TABLE: GENERALIZED LINEAR (MIXED) MODELS IN SAS

	Linear models	**Logistic models**	**Poisson models**
Response variable	**Continuous**	**Binary or polytomous**	**Count variable**
Discrete predictors	Linear ANOVA	Logistic ANOVA	Poisson ANOVA
Continuous predictors	Linear regression	Logistic regression	Poisson regression
Both discrete and continuous predictors	Linear ANCOVA	Logistic ANCOVA	Poisson ANCOVA
(1) SAS procedures assuming iid(*) observations	PROC REG (fixed-effects models)	PROC LOGISTIC (fixed-effects models)	(none)
(2a) Design-based SAS procedures for correlated data	PROC SURVEYREG (fixed-effects models)	PROC SURVEYLOGISTIC (fixed-effects models)	(none)
(2b) Model-based SAS procedures for correlated data	PROC GENMOD PROC MIXED PROC GLIMMIX mixed models	PROC GENMOD PROC GLIMMIX (mixed models)	PROC GENMOD PROC GLIMMIX (mixed models)

*iid = independent identically distributed (corresponds to simple random sampling with replacement, SRSWR)

(Topics in Survey Methodology and Survey Sampling, Risto Lehtonen)

Previous Page | Next Page

SAS model-based procedures

The MIXED Procedure

Overview Getting Started Syntax Details Examples References

Overview: MIXED Procedure

The MIXED procedure fits a variety of mixed linear models to data and enables you to use these fitt
linear model is a generalization of the standard linear model used in the GLM procedure, the gener
nonconstant variability. The mixed linear model, therefore, provides you with the flexibility of modeli
model) but their variances and covariances as well.

The primary assumptions underlying the analyses performed by PROC MIXED are as follows:

- The data are normally distributed (Gaussian).
- The means (expected values) of the data are linear in terms of a certain set of parameters.
- The variances and covariances of the data are in terms of a different set of parameters, and tl
  MIXED.

Since Gaussian data can be modeled entirely in terms of their means and variances/covariances, tl
the complete probability distribution of the data. The parameters of the mean model are referred to
covariance model are referred to as *covariance parameters*.

The fixed-effects parameters are associated with known explanatory variables, as in the standard li
traditional analysis of variance) or quantitative (as in standard linear regression). However, the cova
from the standard linear model.

The need for covariance parameters arises quite frequently in applications, the following being the t

- The experimental units on which the data are measured can be grouped into clusters, and the
- Repeated measurements are taken on the same experimental unit, and these repeated meas

The first scenario can be generalized to include one set of clusters nested within another. For exam
into classes, which in turn can be clustered into schools. Each level of this hierarchy can introduce
scenario occurs in longitudinal studies, where repeated measurements are taken over time. Alterna
nature.

PROC MIXED provides a variety of covariance structures to handle the previous two scenarios. Th
effects parameters, which are additional unknown random variables assumed to affect the variabilit
commonly known as *variance components*, become the covariance parameters for this particular st
random-effects parameters, and, in fact, it is the combination of these two types of effects that led t
traditional variance component models but numerous other covariance structures as well.

PROC MIXED fits the structure you select to the data by using the method of *restricted maximum li*
here that the Gaussian assumption for the data is exploited. Other estimation methods are also ava
behind these estimation methods are discussed in subsequent sections.

After a model has been fit to your data, you can use it to draw statistical inferences via both the fixe
several different statistics suitable for generating hypothesis tests and confidence intervals. The val
covariance model you select, so it is important to choose the model carefully. Some of the output fr
with others.

- **Basic Features**
  Notation for the Mixed Model

Previous Page | Next Page

The GENMOD Procedure

Overview Getting Started Syntax Details Examples References

Overview: GENMOD Procedure

The GENMOD procedure fits generalized linear models, as defined by Nelder and Wedderburn (19 traditional linear models that allows the mean of a population to depend on a *linear predictor* throug distribution to be any member of an exponential family of distributions. Many widely used statistical linear models with normal errors, logistic and probit models for binary data, and log-linear models fo formulated as generalized linear models by the selection of an appropriate link function and respons

See McCullagh and Nelder (1989) for a discussion of statistical modeling using generalized linear n also excellent references with many examples of applications of generalized linear models. Firth (1 Montgomery, and Vining (2002) provide applications of generalized linear models in the engineerin comprehensive accounts of generalized linear models when the responses are binary.

The analysis of correlated data arising from repeated measurements when the measurements are a However, the normality assumption might not always be reasonable; for example, different methodo discrete and correlated. Generalized estimating equations (GEEs) provide a practical method with r

Liang and Zeger (1986) introduced GEEs as a method of dealing with correlated data when, excep a generalized linear model. For example, correlated binary and count data in many cases can be m

The GENMOD procedure can fit models to correlated responses by the GEE method. You can use structures from Liang and Zeger (1986) by using GEEs. See Hardin and Hilbe (2003), Diggle, Liang GEEs.

Bayesian analysis of generalized linear models can be requested by using the BAYES statement in parameters are treated as random variables, and inference about parameters is based on the poste distribution is obtained using Bayes' theorem as the likelihood function of the data weighted with a p knowledge or experience of the likely range of values of the parameters of interest into the analysis use a noninformative prior distribution, and the results of the Bayesian analysis will be very similar t of the posterior distribution is often not feasible, and a Markov chain Monte Carlo method by Gibbs distribution. See Chapter 7, Introduction to Bayesian Analysis Procedures, for an introduction to the Bayesian Analysis: Advantages and Disadvantages for a discussion of the advantages and disadva (2001) for a detailed description of Bayesian analysis.

In a Bayesian analysis, a Gibbs chain of samples from the posterior distribution is generated for the quartiles, HPD and credible intervals, correlation matrix) and convergence diagnostics (autocorrelat and Welch tests; the effective sample size; and Monte Carlo standard errors) are computed for eac matrix of the posterior sample. Trace plots, posterior density plots, and autocorrelation function plot parameter.

The GENMOD procedure enables you to perform exact logistic regression, also called exact conditi called exact conditional Poisson regression, by specifying one or more EXACT statements. You ca parameters. The procedure computes two exact tests: the exact conditional score test and the exac specific parameters and corresponding odds ratios where appropriate. Point estimates, standard er

The GENMOD procedure uses ODS Graphics to create graphs as part of its output. For general inf Graphics Using ODS.

- **What Is a Generalized Linear Model?**
- **Examples of Generalized Linear Models**

Previous Page | Next Page

The GLIMMIX Procedure

Overview Getting Started Syntax Details Examples References

Overview: GLIMMIX Procedure

The GLIMMIX procedure fits statistical models to data with correlations or nonconstant variability and where the response is not necessarily normally distributed. These models are known as generalized linear mixed models (GLMM).

GLMMs, like linear mixed models, assume normal (Gaussian) random effects. Conditional on these random effects, data can have any distribution in the exponential family. The exponential family comprises many of the elementary discrete and continuous distributions. The binary, binomial, Poisson, and negative binomial distributions, for example, are discrete members of this family. The normal, beta, gamma, and chi-square distributions are representatives of the continuous distributions in this family. In the absence of random effects, the GLIMMIX procedure fits generalized linear models (fit by the GENMOD procedure).

GLMMs are useful for the following applications:

- estimating trends in disease rates
- modeling CD4 counts in a clinical trial over time
- modeling the proportion of infected plants on experimental units in a design with randomly selected treatments or randomly selected blocks
- predicting the probability of high ozone levels in counties
- modeling skewed data over time
- analyzing customer preference
- joint modeling of multivariate outcomes

Such data often display correlations among some or all observations as well as nonnormality. The correlations can arise from repeated observation of the same sampling units, shared random effects in an experimental design, spatial (temporal) proximity, multivariate observations, and so on.

The GLIMMIX procedure does not fit hierarchical models with nonnormal random effects. With the GLIMMIX procedure you select the distribution of the response variable conditional on normally distributed random effects.

For more information about the differences between the GLIMMIX procedure and SAS procedures that specialize in certain subsets of the GLMM models, see the section PROC GLIMMIX Contrasted with Other SAS Procedures.

Previous Page | Next Page | Top of Page

Previous Page | Next Page

The GLIMMIX Procedure

Overview Getting Started Syntax Details Examples References

Logistic Regressions with Random Intercepts

PROC GLIMMIX Contrasted with Other SAS Procedures

The GLIMMIX procedure generalizes the MIXED and GENMOD procedures in two important ways. procedure assumes that the response is normally (Gaussian) distributed. Second, the GLIMMIX pro subject-specific (conditional) and population-averaged (marginal) inference. The GENMOD procedu

The GLIMMIX and MIXED procedure are closely related; see the syntax and feature comparison in remainder of this section compares the GLIMMIX procedure with the GENMOD, NLMIXED, LOGIS

The GENMOD procedure fits generalized linear models for independent data by maximum likelihoo approach of Liang and Zeger (1986) and Zeger and Liang (1986). The GEE implementation in the G incorporate random effects. The GEE estimation in the GENMOD procedure relies on R-side covari the method of moments. The GLIMMIX procedure allows G-side random effects and R-side covaria the covariance parameters are not estimated by the method of moments. The parameters are estim GENMOD procedures fit a generalized linear model where the distribution contains a scale parame binomial distribution, the scale parameter is reported in the "Parameter Estimates" table. For some the section Scale and Dispersion Parameters for details about how the GLIMMIX procedure param reported quantities differ between the two procedures.

Many of the fit statistics and tests in the GENMOD procedure are based on the likelihood. In a GLM Even if the log likelihood is tractable, it might be computationally infeasible. In some cases, the obje other cases, only the first two moments of the marginal distribution can be approximated. Consequ many generalized linear mixed models. The GLIMMIX procedure relies heavily on linearization and confidence intervals. Likelihood ratio tests and confidence intervals for covariance parameters are a statement.

The NLMIXED procedure fits nonlinear mixed models where the conditional mean function is a gen models is a special case of the nonlinear mixed models; hence some of the models you can fit with NLMIXED procedure relies by default on approximating the marginal log likelihood through adaptive likelihood estimation by adaptive Gaussian quadrature is available with the METHOD=QUAD optio methods thus differ between the NLMIXED and GLIMMIX procedures, because adaptive quadratur GLIMMIX procedure. If you choose METHOD=LAPLACE or METHOD=QUAD(QPOINTS=1) in the the GLIMMIX procedure performs maximum likelihood estimation based on a Laplace approximatio QPOINTS=1 option in the NLMIXED procedure.

The LOGISTIC and CATMOD procedures also fit generalized linear models; PROC LOGISTIC acc multinomial models for ordered data, and generalized logit models that can be fit with PROC LOGIS tools and capabilities specific to such data implemented in the LOGISTIC procedure go beyond the

Previous Page | Next Page | Top of Page

* Topics in Survey Methodology and Survey Analysis 2013;

```
* SAS data set OHC (Occupational Health Care Survey)
Clustered (Hierarchical, Multilevel) data

Complex sampling design:
Stratified one-stage and two-stage cluster sampling

In analysis phase the data are treated as one-stage cluster sampling
design with workplaces (establishments) as the sample clusters.
This simplifies calculation and is used as the default in SAS,
SPSS and Mplus procedures.

Features of the data set:
H = 5 strata (Industry type and size of workplace)
m = 250 sample clusters (establishments/workplaces)
n = 7841 persons
p = 12 variables

Data are real survey data and have been anonymized and cleaned
for pedagogical purposes (no missing data, weights are constant)

* Methods
Design-based procedures - accounting for clustering effects

SAS SURVEY procedures
Descriptives - SURVEYMEANS

* Access to SAS data library:
- Use the "New library" button
- Use the libname statement;

libname a "...your libref...";

options nocenter;
ods html;

data ohc;
set a.ohc;
run;

proc surveymeans data=ohc nobs mean;
title1 "OHC Survey data";
title2 "Design-based analysis";
var psych psych2 phys chron;     * Variable list;
domain sex;                      * Subgroup analysis;
strata stratum;                  * Stratum variable;
cluster psu;                     * Cluster variable;
run;
```

The SURVEYMEANS Procedure

Data Summary	
Number of Strata	5
Number of Clusters	250
Number of Observations	7841

Statistics

Variable	Label	N	Mean	Std Error of Mean
PSYCH	Psychic strain - 1st princomp	7841	-2.46015E-11	0.015839
PSYCH2	Psychic strain - dichotomy	7841	0.499426	0.007336
PHYS	Physical health hazards of work	7841	0.345747	0.014385
CHRON	Chronic morbidity	7841	0.292437	0.006808

Domain Analysis: Gender

Gender	Variable	Label	N	Mean	Std Error of Mean
1	**PSYCH**	Psychic strain - 1st princomp	4485	-0.100784	0.017953
	PSYCH2	Psychic strain - dichotomy	4485	0.454849	0.008773
	PHYS	Physical health hazards of work	4485	0.459532	0.016659
	CHRON	Chronic morbidity	4485	0.292977	0.009199
2	**PSYCH**	Psychic strain - 1st princomp	3356	0.134689	0.024198
	PSYCH2	Psychic strain - dichotomy	3356	0.558999	0.011008
	PHYS	Physical health hazards of work	3356	0.193683	0.013966
	CHRON	Chronic morbidity	3356	0.291716	0.009710

Topics in Social Statistics, fall 2013

Logistic ANCOVA for cluster correlated data
SAS example for OCH Survey data set

```
libname a " SAS data library reference";

proc surveylogistic data=a.ohc;
title "OHC data / SURVEYLOGISTIC";
title2 "Design-based logistic ANCOVA";
strata stratum;
cluster PSU;
class sex / param=ref;
model psych2(event=last)=sex age phys chron  sex*age
     / link=logit clodds rsquare ; run;
```

```
The SURVEYLOGISTIC Procedure

                              Model Information

Data Set                      A.OHC
Response Variable             PSYCH2                     Psychic strain - dichotomy
Number of Response Levels     2
Stratum Variable              STRATUM                    Stratum identifier
Number of Strata              5
Cluster Variable              PSU
Number of Clusters            250
Model                         Binary Logit
Optimization Technique        Fisher's Scoring
Variance Adjustment           Degrees of Freedom (DF)


            Variance Estimation

Method                           Taylor Series
Variance Adjustment    Degrees of Freedom (DF)


Number of Observations Read        7841
Number of Observations Used        7841


          Response Profile

 Ordered                      Total
   Value       PSYCH2     Frequency

       1            0          3925
       2            1          3916

Probability modeled is PSYCH2=1.


   Class Level Information

                     Design
Class     Value     Variables

SEX       1                 1
          2                 0

R-Square    0.0326    Max-rescaled R-Square    0.0434
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	259.7153	5	<.0001
Score	256.4213	5	<.0001
Wald	203.3968	5	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
SEX	1	23.8111	<.0001
AGE	1	1.2611	0.2614
PHYS	1	21.5044	<.0001
CHRON	1	96.2804	<.0001
AGE*SEX	1	6.5458	0.0105

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.1964	0.1573	1.5600	0.2117
SEX	1	1	-0.9925	0.2034	23.8111	<.0001
AGE		1	-0.00456	0.00406	1.2611	0.2614
PHYS		1	0.2764	0.0596	21.5044	<.0001
CHRON		1	0.5641	0.0575	96.2804	<.0001
AGE*SEX	1	1	0.0131	0.00511	6.5458	0.0105

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PHYS	1.318	1.173	1.482
CHRON	1.758	1.571	1.967

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.0	Somers' D	0.209
Percent Discordant	39.2	Gamma	0.210
Percent Tied	0.8	Tau-a	0.104
Pairs	15370300	c	0.604

Wald Confidence Interval for Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
PHYS	1.0000	1.318	1.173	1.482
CHRON	1.0000	1.758	1.571	1.967

* Topics in Survey Methodology and Survey Analysis 2013;

```
* Topics in Survey Methodology and Survey Analysis 2012;
* Design-based and model-based analysis of complex survey data;
* PART 1: Descriptives and simple tests;

* SAS data set OHC (Occupational Health Care Survey)
Clustered (Hierarchical, Multilevel) data

Complex sampling design: Stratified one-stage and two-stage cluster sampling

In analysis phase the data are treated as one-stage cluster sampling
design with workplaces (establishments) as the sample clusters.
This simplifies calculation and is used as the default in SAS,
SPSS and Mplus procedures.

Features of the data set:
H = 5 strata (Industry type and size of workplace)
m = 250 sample clusters (establishments/workplaces)
n = 7841 persons
p = 12 variables

Data are real survey data and have been anonymized and cleaned
for pedagocical purposes (no missing data, weights are constant)

SPSS use: CSPLAN file (sample plan data set) will be created
in PC session
;

* Methods

(1) Design-based procedures - accounting for clustering effects

SAS SURVEY design-based procedures
  Descriptives: SURVEYMEANS
  Test of independence: SURVEYFREQ
  Logistic regression: SURVEYLOGISTIC
SPSS Complex Samples module, design-based procedures
  CSPLAN - Complex samples plan
  DESCRIPTIVES - Means, proportions etc.
  CROSSTABS - Frequency tables and tests of independence
  CSLOGISTIC - Logistic regression
Mplus (COMPLEX, TWOLEVEL)
  Logistic regression

(2) Model-based procedures, hierarchical (multilevel) analysis

SAS Logistic regression
  GENMOD (GEE/Exchangeable estimation)
  GLIMMIX (Generalixed linear mixed modelling)
SPSS
  GENERALIZED LINEAR MODELS - Generalized estimating equations GEE
  MIXED MODELS - Generalized linear mixed models

NOTE: See also VLISS Training Key #298
;
```

```
options nocenter;

* Access to SAS data library:
- Use the "New library" button
- Use the libname statement;

*libname a "Z:\Documents\My SAS Files\9.3\Social Statistics Course 2013";
libname a "I:\Root\USB\HY\Social Statistics Course\Course 2013\SAS Data";

data ohc;
set a.ohc;
run;

* see HELP proc contents;
proc contents data=ohc varnum;
title1 "OHC Survey";
title2 "Variable list";
run;

* see HELP proc surveymeans;
proc surveymeans data=ohc nobs mean;
title1 "OHC Survey";
title2 "Design-based analysis: Means by gender";
var psych psych2 phys chron;
domain sex;
strata stratum;
cluster psu;
run;

* Let us carry out the same analysis using SPSS;

* see HELP proc surveyfreq;
proc surveyfreq data=ohc;
title1 "OHC Survey";
title2 "Design-based analysis: Frequency table and chi-square test";
tables phys*psych2 / chisq cl;
strata stratum;
cluster psu;
run;

* Let us carry out the same analysis using SPSS;
```

```
* Topics in Survey Methodology and Survey Analysis 2013;

* Design-based and model-based analysis of complex survey data;
* Analysis of cluster correlated data;

* SAS data set OHC (Occupational Health Care Survey)
Clustered (Hierarchical, Multilevel) data

Complex sampling design:
Stratified one-stage and two-stage cluster sampling

In analysis phase the data are treated as one-stage cluster sampling design with
workplaces (establishments) as the sample clusters. This simplifies calculation and
is used as the default in SAS, SPSS and Mplus.

SAS:
Design-based analysis:
Procedure SURVEYFREQ (design-based Rao-Scott corrected tests for independence)
Procedure SURVEYLOGISTIC (Pseudo maximum likelihood estimation)
Model-based analysis (Multilevel modelling):
Procedure GENMOD (GEE estimation /Exchangeable correlation structure)
Procedure GLIMMIX (Generalized linear mixed modelling)

SPSS:
Complex Samples module - Design-based analysis
CSPLAN file (sample plan data set) will be created in PC session
CSLOGISTIC for Logistic regression

Mplus: Design-based analysis
COMPLEX, TWOLEVEL
Logistic regression

NOTE: See also VLISS Training Key #298

SAS code will be worked out further during PC session.
;


* Access to SAS data library: Use "New Library" button;


options nocenter;

data ohc;
set a.ohc;
run;
```

```
* LOGISTIC ANCOVA;

* SURVEYLOGISTIC;
* Design-based analysis;
* class variable sex
Parametrization:
Effect coding (default)
Reference cell coding (dummy coding) - useful for odds ratios!
;
proc surveylogistic data=ohc;
title "OHC data / SURVEYLOGISTIC";
title2 "Design-based logistic ANCOVA";
strata stratum;
cluster PSU;
* Effect coding;
*class sex;
* Reference cell coding (reference class=females);
*class sex(ref="2") / param=ref;
* Reference cell coding (reference class=males);
class sex(ref="1")/ param=ref ;
* Note: event=last means that the probability modelled = Pr(psych2=1);
model psych2(event=last)=sex age phys chron sex*age
          / link=logit clodds rsquare;
run;

* Multilevel modelling;
* GENMOD - GEE method;
* Model-based analysis;
proc genmod data=ohc descending;
title "OHC data / Logistic ANCOVA";
title2 "Model-based analysis";
title3 "PROC GENMOD, GEE/Exchangeable";
class sex(ref=first) PSU;
model psych2=sex age phys chron sex*age
          / dist=bin link=logit;
repeated subject=PSU / type=exch;
run;

* GLIMMIX;
* Model-based analysis - hierarchical (multilevel) model;
proc glimmix data=ohc empirical;
title "OHC data / Logistic ANCOVA";
title2 "Model-based analysis";
title3 "PROC GLIMMIX, logistic mixed model";
class sex PSU;
model psych2(event=last)=sex age phys chron sex*age
          / dist=bin link=logit solution oddsratio;
random int / subject=PSU type=vc ;
run;

* Let us carry out this analysis by SPSS and Mplus;
```

SPSS CSPLAN file for OHC Survey

Code:

```
GET
  FILE='...\SPSS\ohc.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
* Analysis Preparation Wizard.
CSPLAN ANALYSIS
  /PLAN FILE='...\OHC.csaplan'
  /PLANVARS ANALYSISWEIGHT=wd
  /SRSESTIMATOR TYPE=WOR
  /PRINT PLAN
  /DESIGN STRATA=STRATUM CLUSTER=PSU
  /ESTIMATOR TYPE=WR.
```

SPSS Output:

Complex Samples: Plan

[DataSet1] I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\ohc.sav

Warnings

This procedure does not check the consistency of the working data file with the plan file. We recommend looking at the output table or the plan file to check consistency before performing selection or analysis.

Summary

			Stage 1
Design Variables	Stratification	1	Stratum identifier
	Cluster	1	Primary sampling unit (Cluster)
Analysis Information	Estimator Assumption		Sampling with replacement

Plan File: I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\OHC.csaplan

Weight Variable: Design weight

SRS Estimator: Sampling without replacement

SPSS - Subgroup Means
under cluster correlated OHC data

```
* Complex Samples Descriptives.
CSDESCRIPTIVES
  /PLAN FILE='I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\OHC.csaplan'
  /SUMMARY VARIABLES=PSYCH PSYCH2 PHYS CHRON
  /SUBPOP TABLE=SEX DISPLAY=LAYERED
  /MEAN
  /STATISTICS SE COUNT DEFF
  /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

Complex Samples: Descriptives

```
[DataSet1] I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\ohc.sav
```

Univariate Statistics

		Estimate	Standard Error	Design Effect	Unweighted Count
Mean	Psychic strain - 1st princomp	,00	,016	1,976	7841
	Psychic strain - dichotomy	,50	,007	1,695	7841
	Physical health hazards of work	,35	,014	7,203	7841
	Chronic morbidity	,29	,007	1,764	7841

Subpopulation Descriptives

Univariate Statistics

Gender			Estimate	Standard Error	Design Effect
1	Mean	Psychic strain - 1st princomp	-,10	,018	1,566
		Psychic strain - dichotomy	,45	,009	1,398
		Physical health hazards of work	,46	,017	5,033
		Chronic morbidity	,29	,009	1,840
2	Mean	Psychic strain - 1st princomp	,13	,024	1,852
		Psychic strain - dichotomy	,56	,011	1,657
		Physical health hazards of work	,19	,014	4,209
		Chronic morbidity	,29	,010	1,538

Univariate Statistics

Gender			Unweighted Count
1	Mean	Psychic strain - 1st princomp	4485
		Psychic strain - dichotomy	4485
		Physical health hazards of work	4485
		Chronic morbidity	4485
2	Mean	Psychic strain - 1st princomp	3356
		Psychic strain - dichotomy	3356
		Physical health hazards of work	3356
		Chronic morbidity	3356

SPSS - Test of independence in two-way table under cluster correlated OHC data

```
* Complex Samples Crosstabs.
CSTABULATE
  /PLAN FILE='I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\OHC.csaplan'
  /TABLES VARIABLES=PHYS BY PSYCH2
  /CELLS COLPCT
  /STATISTICS SE DEFF
  /TEST INDEPENDENCE
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Complex Samples: Tables

```
[DataSet1] I:\Root\USB\HY\Social Statistics Course\Course 2013\SPSS\ohc.sav
```

Physical health hazards of work * Psychic strain - dichotomy

Physical health hazards of work			Psychic strain - dichotomy		
			0	1	Total
0	% within Psychic strain - dichotomy	Estimate	67,0%	63,9%	65,4%
		Standard Error	1,5%	1,6%	1,4%
		Design Effect	4,065	4,548	7,203
1	% within Psychic strain - dichotomy	Estimate	33,0%	36,1%	34,6%
		Standard Error	1,5%	1,6%	1,4%
		Design Effect	4,065	4,548	7,203
Total	% within Psychic strain - dichotomy	Estimate	100,0%	100,0%	100,0%
		Standard Error	0,0%	0,0%	0,0%
		Design Effect	.	.	.

Tests of Independence

		Chi-Square	Adjusted F	df1	df2
Physical health hazards of work * Psychic strain - dichotomy	Pearson	8,407	6,171	1	245
	Likelihood Ratio	8,409	6,172	1	245

Tests of Independence

		Sig.
Physical health hazards of work * Psychic strain - dichotomy	Pearson	,014
	Likelihood Ratio	,014

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

Topics in Social Statistics, fall 2013
Logistic ANCOVA for cluster correlated data
SPSS example for OCH Survey data set

```
GET
  FILE='...your fileref ..\Data\OHC.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
* Analysis Preparation Wizard.
CSPLAN ANALYSIS
 /PLAN FILE='E:\Root\USB\HY\Social Statistics Course '+
  '2010\Data\OHC.csaplan'
 /PLANVARS ANALYSISWEIGHT=w
 /SRSESTIMATOR TYPE=WOR
 /PRINT PLAN
 /DESIGN  STRATA= STRATUM  CLUSTER= PSU
   /ESTIMATOR TYPE=WR.
```

Complex Samples: Plan

```
[DataSet1]  ... \OHC.sav
```

Warnings

This procedure does not check the consistency of the working data file with the plan file. We recommend looking at the output table or the plan file to check consistency before performing selection or analysis.

Summary

			Stage 1
Design Variables	Stratification	1	Stratum identifier
	Cluster	1	PSU
Analysis Information	Estimator Assumption		Sampling with replacement

Plan File: E:\Root\USB\HY\Social Statistics Course 2010\Data\OHC.csaplan
Weight Variable: w
SRS Estimator: Sampling without replacement

```
* Complex Samples Logistic Regression.
CSLOGISTIC PSYCH2(LOW) BY SEX WITH AGE PHYS CHRON
 /PLAN FILE = ' ....\OHC.csaplan'
 /MODEL SEX AGE PHYS CHRON SEX*AGE
 /INTERCEPT INCLUDE=YES SHOW=YES
 /STATISTICS PARAMETER EXP SE CINTERVAL TTEST DEFF
 /TEST TYPE=F PADJUST=LSD
 /ODDSRATIOS COVARIATE=[PHYS(1)]
 /ODDSRATIOS COVARIATE=[CHRON(1)]
 /MISSING CLASSMISSING=EXCLUDE
 /CRITERIA MXITER=100 MXSTEP=5
  PCONVERGE=[1e-006 RELATIVE] LCONVERGE=[0] CHKSEP=20
  CILEVEL=95
 /PRINT SUMMARY VARIABLEINFO SAMPLEINFO.
```

Complex Samples: Logistic Regression

[... \OHC.sav

Warnings

The estimated parameter variance for the simple random sampling design is zero. The design effects, adjusted Wald F test, or adjusted Wald Chi-square test cannot b computed.

Sample Design Information

		N
Unweighted Cases	Valid	7841
	Invalid	0
	Total	7841
Population Size		7841,000
Stage 1	Strata	5
	Units	250
Sampling Design Degrees of Freedom		245

Categorical Variable Information

		Weighted Count	Weighted Percent
Psychic strain - dichotomy[b]	0[a]	3925,000	50,1%
	1	3916,000	49,9%
Gender	1	4485,000	57,2%
	2	3356,000	42,8%
Population Size		7841,000	100,0%

a. Reference Category

b. Dependent Variable

Covariate Information

	Mean
Age in years	37,58
Physical health hazards of work	,35
Chronic morbidity	,29

Pseudo R Squares

Cox and Snell	,033
Nagelkerke	,043
McFadden	,024

Dependent Variable: Psychic strain - dichotomy
(reference category = 0)
Model: (Intercept), SEX, AGE, PHYS, CHRON, SEX * AG

Tests of Model Effects

Source	df1	df2	Wald F	Sig.
(Corrected Model)	5,000	241,000	40,044	,000
(Intercept)	1,000	245,000	8,591	,004
SEX	1,000	245,000	23,828	,000
AGE	1,000	245,000	,565	,453
PHYS	1,000	245,000	21,521	,000
CHRON	1,000	245,000	96,351	,000
SEX * AGE	1,000	245,000	6,550	,011

Dependent Variable: Psychic strain - dichotomy (reference category = 0)
Model: (Intercept), SEX, AGE, PHYS, CHRON, SEX * AGE

Parameter Estimates

				95% Confidence Interval		Hypothesis Test					95% Confidence Interval for Exp(B)	
Psychic strain - dichotomy	Parameter	B	Std. Error	Lower	Upper	t	df	Sig.	Design Effect	Exp(B)	Lower	Upper
1	(Intercept)	,196	,157	-,113	,506	1,249	245,000	,213	.	1,217	,893	1,659
	[SEX=1]	-,993	,203	-1,393	-,592	-4,881	245,000	,000	.	,371	,248	,553
	[SEX=2]	,000[a]	.	.	.	.	.	.	.	1,000	.	.
	AGE	-,005	,004	-,013	,003	-1,123	245,000	,262	.	,995	,988	1,003
	PHYS	,276	,060	,159	,394	4,639	245,000	,000	.	1,318	1,172	1,483
	CHRON	,564	,057	,451	,677	9,816	245,000	,000	.	1,758	1,570	1,969
	[SEX=1] * AGE	,013	,005	,003	,023	2,559	245,000	,011	.	1,013	1,003	1,023
	[SEX=2] * AGE	,000[a]	.	.	.	.	.	.	.	1,000	.	.

Dependent Variable: Psychic strain - dichotomy (reference category = 0)
Model: (Intercept), SEX, AGE, PHYS, CHRON, SEX * AGE

a. Set to zero because this parameter is redundant.

Odds Ratios 1[a]

Units of Change		Psychic strain - dichotomy	Odds Ratio	95% Confidence Interval	
				Lower	Upper
Physical health hazards of work	1,000	1	1,318	1,172	1,483

Dependent Variable: Psychic strain - dichotomy (reference category = 0)
Model: (Intercept), SEX, AGE, PHYS, CHRON, SEX * AGE

a. Factors and covariates used in the computation are fixed at the following values: Gender=2; Age in years=37,58; Physical health hazards of work=,35; Chronic morbidity=,29

Odds Ratios 2[a]

Units of Change		Psychic strain - dichotomy	Odds Ratio	95% Confidence Interval	
				Lower	Upper
Chronic morbidity	1,000	1	1,758	1,570	1,969

Dependent Variable: Psychic strain - dichotomy (reference category = 0)
Model: (Intercept), SEX, AGE, PHYS, CHRON, SEX * AGE

a. Factors and covariates used in the computation are fixed at the following values: Gender=2; Age in years=37,58; Physical health hazards of work=,35; Chronic morbidity=,29

Mplus VERSION 5.21 LOGISTIC REGRESSION 20 Sept. 2013 (Risto Lehtonen)
MUTHEN & MUTHEN

```
INPUT INSTRUCTIONS

  TITLE:
  Mplus Logistic Regression for OHC Survey data;
  COMPLEX type analysis for cluster correlated data;

  DATA:
  FILE IS
  "Z:\Documents\My SAS Files\9.2\Social Statistics Course 2011\OHC.inp";
  TYPE IS INDIVIDUAL;

  VARIABLE:
    NAMES ARE ID STRATUM SEX AGE AGE2 PHYS CHRON PSYCH2 PSU;
    USEVARIABLES ARE STRATUM PSU PSYCH2 AGE PHYS CHRON SEX01 INT;
    CATEGORICAL IS PSYCH2;
    STRATIFICATION IS STRATUM;
    CLUSTER IS PSU;

  DEFINE:
    SEX01=SEX-1;
    INT=SEX01*AGE;

  ANALYSIS:
    TYPE IS COMPLEX;
    ESTIMATOR IS MLR;
    LINK IS LOGIT;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

  MODEL:
     PSYCH2 ON SEX01 AGE PHYS CHRON INT;

  OUTPUT:  SAMPSTAT CINTERVAL;


INPUT READING TERMINATED NORMALLY

Mplus Logistic Regression for OHC Survey data;
COMPLEX type analysis for cluster correlated data;

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                        7841

Number of dependent variables                                    1
Number of independent variables                                  5
Number of continuous latent variables                            0

Observed dependent variables

  Binary and ordered categorical (ordinal)
   PSYCH2

Observed independent variables
   AGE         PHYS        CHRON       SEX01       INT

Variables with special functions

  Stratification        STRATUM
  Cluster variable      PSU
```

```
Estimator                                                       MLR
Information matrix                                         OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations                                 1000
  Convergence criterion                                  0.500D-04
Optimization Specifications for the EM Algorithm
  Maximum number of iterations                                  500
  Convergence criteria
    Loglikelihood change                                 0.100D-02
    Relative loglikelihood change                        0.100D-05
    Derivative                                           0.100D-02
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
  Number of M step iterations                                     1
  M step convergence criterion                           0.100D-02
  Basis for M step termination                           ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
  Number of M step iterations                                     1
  M step convergence criterion                           0.100D-02
  Basis for M step termination                           ITERATION
  Maximum value for logit thresholds                            15
  Minimum value for logit thresholds                           -15
  Minimum expected cell size for chi-square              0.100D-01
Optimization algorithm                                         EMA
Integration Specifications
  Type                                                    STANDARD
  Number of integration points                                  15
  Dimensions of numerical integration                            0
  Adaptive quadrature                                           ON
Link                                                         LOGIT
Cholesky                                                       OFF

Input data file(s)
  Z:\Documents\My SAS Files\9.2\Social Statistics Course 2011\OHC.inp
Input data format  FREE


SUMMARY OF DATA

     Number of strata            5
     Number of clusters        250


SUMMARY OF CATEGORICAL DATA PROPORTIONS

    PSYCH2
      Category 1    0.501
      Category 2    0.499
```

SAMPLE STATISTICS

SAMPLE STATISTICS

Means

	AGE	PHYS	CHRON	SEX01	INT
1	37.582	0.346	0.292	0.428	16.112

Covariances

	AGE	PHYS	CHRON	SEX01	INT
AGE	114.272				
PHYS	-0.202	0.226			
CHRON	1.374	0.009	0.207		
SEX01	0.027	-0.065	0.000	0.245	
INT	51.966	-2.385	0.598	9.217	397.936

Correlations

	AGE	PHYS	CHRON	SEX01	INT
AGE	1.000				
PHYS	-0.040	1.000			
CHRON	0.283	0.042	1.000		
SEX01	0.005	-0.277	-0.001	1.000	
INT	0.244	-0.251	0.066	0.934	1.000

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Loglikelihood

H0 Value	-5305.104
H0 Scaling Correction Factor for MLR	1.383

Information Criteria

Number of Free Parameters	6
Akaike (AIC)	10622.208
Bayesian (BIC)	10664.011
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	10644.944

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
PSYCH2	ON				
SEX01		0.993	0.203	4.881	0.000
AGE		0.009	0.003	2.633	0.008
PHYS		0.276	0.060	4.639	0.000
CHRON		0.564	0.057	9.816	0.000
INT		-0.013	0.005	-2.559	0.010
Thresholds					
PSYCH2$1		0.796	0.130	6.127	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

PSYCH2 ON	
SEX01	2.698
AGE	1.009
PHYS	1.318
CHRON	1.758
INT	0.987

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix 0.752E-05
(ratio of smallest to largest eigenvalue)

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower .5%	Lower 2.5%	Estimate	Upper 2.5%	Upper .5%
PSYCH2 ON					
SEX01	0.469	0.594	0.993	1.391	1.516
AGE	0.000	0.002	0.009	0.015	0.017
PHYS	0.123	0.160	0.276	0.393	0.430
CHRON	0.416	0.451	0.564	0.677	0.712
INT	-0.026	-0.023	-0.013	-0.003	0.000
Thresholds					
PSYCH2$1	0.461	0.541	0.796	1.051	1.131

CONFIDENCE INTERVALS FOR THE LOGISTIC REGRESSION ODDS RATIO RESULTS

PSYCH2 ON					
SEX01	1.598	1.811	2.698	4.019	4.555
AGE	1.000	1.002	1.009	1.015	1.017
PHYS	1.131	1.173	1.318	1.482	1.537
CHRON	1.516	1.571	1.758	1.967	2.038
INT	0.974	0.977	0.987	0.997	1.000

Beginning Time: 10:46:58
Ending Time: 10:46:58
Elapsed Time: 00:00:00

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA 90066

Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com