# Software for GEE: PROC GENMOD and SUDAAN

Babubhai V. Shah, Research Triangle Institute, Research Triangle Park, NC

## Abstract

Until recently, most of the statistical software was limited to analyzing data from simple random samples. Recently, some programs have become available to analyze correlated or clustered data. The major advantage of these methods is the use of robust sandwich estimators for the variances of parameter estimates. We shall review two such programs and discuss their strong points. We shall also present pointers for taking greatest advantage of both these programs.

## Introduction

SAS/STAT™ is primarily designed to analyze data from a simple random sample where all observations are independent of each other. SAS has two new procedures GENMOD and MIXED that analyze clustered or (correlated) observations. SUDAAN™ is designed specifically to analyze clustered (or correlated) data. All the procedures in SUDAAN provide robust estimators of variances for all the estimates.

This paper compares only GEE capabilities in the two packages, specifically, SAS GENMOD procedure and several modeling procedures in SUDAAN.

The primary objective of this paper is to consider the similarities and differences in the available software. We shall review the theory and assumptions underlying generalized estimating equations (GEE), and consider various choices you have in theory to analyze your data. We shall compare SAS and SUDAAN implementations with respect to:

- Link function for mean,
- Mean variance relationship,
- Variance estimation method
- tests of hypotheses.

You can apply this information to determine the software that will be most appropriate for your specific problems. Of course, as new and improved releases of software become available, your choices will have to be adjusted in future.

## An Example of Logistic Regression

We shall start with a simple example of the type of problem to illustrate the issues before considering the theory underlying GEE. Consider a teratology study of the malformations due to an exposure to a toxic chemical. The observed outcome y is a binary variable representing the presence or absence of a malformation. The explanatory variable x is the exposure level or dose. The logistic model for the probability of malformation in the ith observation is

$$E(y_i) = \mu_i, \quad \ln(\mu_i) - \ln(1-\mu_i) = \beta_0 + \beta_1 x_i \qquad (1)$$

Of course, the outcome for the fetuses from the same litter are correlated with each other, and we cannot assume that the all the observations are independent. The variance of y is

$$V(y_i) = \mu_i (1 - \mu_i).$$

The correlation between $y_i$ and $y_j$ is equal to ? if $y_i$ and $y_j$ belong to the same litter, and is equal to zero otherwise.

The features of this example that you need to focus on are:

(1) The relationship in Equation (1) between $\mu$ and $\beta = [\beta_0, \beta_1]$ is not linear, but is a function of $L = x\beta$.
(2) The variance covariance matrix of $y = [y_1, y_2, ..., y_n]$ is known and is a function of the means.
(3) The first two moments of $y$ are specified, but it is not possible to write down the likelihood function.

We cannot use the maximum likelihood method of estimation, and need to use a method that is applicable under the assumptions for only two moments.

## Quasi -likelihood functions

In further discussion, we shall consider the general case, where the functions for the mean $\mu$ and variance $V$ are arbitrary. If the function V is constant, the solution based on the weighted least squares is given by the equations

$$U(\beta) = \frac{\partial \mu}{\partial \beta} V^{-1}(y-\mu) = 0. \qquad (2)$$

We apply the same equations, but in this case both $V$ and $\mu$ are functions of $\beta$. The derivative of $U(\beta)$ is

$$J_0 = \frac{\partial U(\beta)}{\partial \beta} = \frac{\partial^2 \mu}{\partial \beta^2} V^{-1}(y-\mu) + \frac{\partial \mu}{\partial \beta} \frac{\partial V^{-1}}{\partial \beta}(y-\mu) - \frac{\partial \mu}{\partial \beta} V^{-1} \frac{\partial \mu}{\partial \beta}. \qquad (3)$$

The following results for the mean and variance of $U(\beta)$, and the expected value of $J_0$ are

$$E[U(\beta)] = 0$$
$$E\left[\frac{\partial U(\beta)}{\partial \beta}\right] = -\frac{\partial \mu}{\partial \beta} V^{-1} \frac{\partial \mu}{\partial \beta}$$
$$Var[U(\beta)] = \sigma^2 \frac{\partial \mu}{\partial \beta} V^{-1} \frac{\partial \mu}{\partial \beta}. \qquad (4)$$

Hence, $U(\beta)$ has same properties as those of the derivative of a log-likelihood function.

If $\hat{\beta}$ is the solution of Equation (2), then

a. The solution $\hat{\beta}$ is consistent, that is it tends to the true value of $\beta$ (in probability) as sample size tends to infinity.
b. The asymptotic distribution of the solution $\hat{\beta}$ is multivariate normal.
c. The variance covariance matrix of the solution $\hat{\beta}$ is

$$Var(\hat{\beta}) = \sigma^2 I_0, \qquad (5)$$

where

$$I_0 = \left[\frac{\partial \mu}{\partial \beta} V^{-1} \frac{\partial \mu}{\partial \beta}\right]^{-1}.$$

For additional information on quasi-likelihood functions, please refer to Nelder and Wedderburn (1972),McCullagh (1983), and McCullagh and Nelder (1989).

The solution $\hat{\beta}$ is consistent, even when the variance assumptions are not satisfied; for example, within cluster correlation is not the same from one pair of observations to another. However, variance of $\hat{\beta}$ is not robust against specification errors in assumptions about $V$. We shall consider several alternatives for computing robust estimates of the variance covariance matrix of $\hat{\beta}$.

## Robust estimation of variance

A robust estimator for the variance of $\hat{\beta}$ was used by survey researchers in specific cases, by Kish and Frankel (1974), Folsom (1976), and Shah, Holt and Folsom (1977). Binder (1983) presents general theory for estimates based on finite populations, for which no distributional assumption are feasible.

Binder's (1983) derivation is based on the Taylor series approximation of the estimate $\hat{\beta}$, which can be written as

$$\hat{\beta} \approx \beta_T - J_0^{-1} \frac{\partial \mu}{\partial \beta} V^{-1}(y-\mu).$$

Hence, the variance of $\hat{\beta}$ Is given by

$$Var_B(\hat{\beta}) \approx J_0^{-1} I_1 [J_0]^{-1}, \quad (6)$$

where

$$I_1 = \frac{\partial \mu}{\partial \beta} V^{-1} Var(y) V^{-1} \frac{\partial \mu}{\partial \beta},$$

where $Var(y)$ is based on the sample data and not on any assumption about $V$, except independence of clusters, such as litters of animals.

If you replace $J_0$ by its expected value $I_0$, you obtain

$$Var_Z(\hat{\beta}) \approx I_0^{-1} I_1 I_0^{-1}. \quad (7)$$

The references mentioned above considered only independent working correlations. Zeger and Liang (1986) presented the results with working correlations and named the method as "Generalized Estimating Equations (GEE)." Zeger and Liang also suggested a simple heuristic method for estimating correlations from the data.

Although Binder's and Zeger's methods are similar, until further research is carried out, we will not know which method is better and under what circumstances. The proponents of each of the methods have their heuristic opinions.

"Zeger's method uses $I_0$ which is the expected value of $J_0$, and is likely to produce more stable estimate of the variance than Binder's method."

"The expected value of $J_0$ is $I_0$ under the assumed model and conditional on the values of the explanatory variables $x$. Binder's method does depend on such conditions and is likely to be more robust than Zeger's method. Further more, Binder's method is more general and valid for set of estimating equations that do not have the same form as Equation (2). For example, in the case of the partial likelihood equation for Cox's proportional hazard model, Zeger's method is not applicable, and you have to use Binder's (1992) method."

In the absence of any compelling evidence, the choice of the method is left to your own preference. You may also wish to look at bootstrap or replication methods, such as "Jackknife" and "Balanced Repeated Replication (BRR)."

## Variance estimation by replication methods

There are several possible replication methods that may be used for robust estimation of variance. The most well known ones are Jackknife and BRR.

Assume that we have H strata with $n_h$ clusters (or primary sampling units --PSUs) in the hth stratum, for h=1, 2, ...,H. Let $\hat{\beta}_{hi}$ be the estimate of the parameter $\beta$ after excluding the data from the ith cluster in the hth stratum. The Jackknife variance estimator of $\hat{\beta}$ obtained from overall sample, is

$$Var_J(\hat{\beta}) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{\beta}_{hi} - \hat{\beta})(\hat{\beta}_{hi} - \hat{\beta})'. \quad (8)$$

This method is known as the "delete one" Jackknife method.

In the BRR method half-replicates are constructed by selecting half the clusters. If we have H strata with 2 clusters in each stratum, then we can select only one cluster from each stratum to create a half replicate. Counting all possible combinations we shall have $2^H$ half replicates. Alternatively if we have 2H clusters with no stratification, then by selecting any H clusters we can create $\binom{2H}{H}$ half replicates. In practice, to keep the task manageable a few, say R $(32 \le R \le 64)$, are created using an orthogonal matrix. The BRR variance of $\hat{\beta}$ is

$$Var_R(\hat{\beta}) = \frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_r - \hat{\beta})(\hat{\beta}_r - \hat{\beta})', \quad (9)$$

where $\hat{\beta}_r$ is the estimate of $\beta$ from the rth half replicate.

For a review of these methods, you may refer to Rao, Wu, and Yue (1992) and Krewski, D. and Rao, J.N.K (1981). These methods are useful when software for estimation is available but does not provide robust estimates of the variance.

## Relevant Features

With the above review of theory, we conclude that the Basic features for GEE software are:

A. Consistent estimates $\hat{\beta}$ of the parameters $\beta$, based on the model for the mean (link function) and the function representing variance, even when the distributional assumptions are not accurate.
B. Robust or consistent estimates of the variance covariance matrix of $\hat{\beta}$, even when mean variance relationship or the correlation structure does not hold for each of the clusters.
C. Permit specification of the most appropriate correlation matrix for observations within a cluster.

There are many other aspects of software, such as input, output, graphics, user friendly interface etc, which you may consider relevant. However, if any software does not meet the basic needs, you should not use it to analyze your data. Furthermore, some of the other aspects may be taken care of by other software tools, once the basic analysis is carried out.

I know of four software packages that meet above basic needs to some extent. These are:

- C SAS PROC GENMOD
- C STATA
- C SUDAAN
- C WESVAR.

We compare only SAS and SUDAAN in this paper. SAS assumes simple random sampling in all of its procedures, except GENMOD and MIXMOD. SUDAAN is designed primarily to analyze clustered or correlated data and is available in a SAS callable version.

Of course, none of the package addresses:

- C all possible models
- C all methods of robust variance estimation
- C all possible correlation structures.

You should evaluate the extent to which a particular software package meets your need for analyzing your data set. We shall discuss the extent to which basic features are covered in SAS PROC GENMOD and SUDAAN.

## Available Models

SAS procedure GENMOD is based on theory for generalized linear model (GLM) and allows you to specify various link and variance functions for univariate outcome variables. SUDAAN has a separate procedure for each model and you need to use a different procedure for each model, and has two procedures that cannot be specified by a link function in GLM.

Some models are common to both and some are in only one of the package. Table 1 presents the most commonly used models.
Both GENMOD and SUDAAN compute robust estimates of variances for some models: linear regression and logistic regression. GENMOD permits Poisson regression, SUDAAN will add a log-linear procedure for count data (LOGLINK) in the next version. GENMOD has a built in facility for probit, gamma, inverse Gaussian and other models. Where as, SUDAAN has procedures for multinomial logistic model for ordinal as well as nominal outcomes, and Cox's proportional hazard model for survival.

| Table 1. Some models for GEE | | |
|---|---|---|
| Model | Link $g(\mu)$ | Variance $V(\mu)$ |
| Linear | $\mu$ | 1 |
| Logit | $\ln(\mu)\&\ln(1\&\mu)$ | $\mu(1\&\mu)$ |
| Probit | $F^{\&1}(\mu)$, F is cumulative normal | $\mu(1\&\mu)$ |
| Poisson | $\ln(\mu)$ | $\mu$ |
| Gamma | $\ln(\mu)$ | $\mu^2$ |
| Multinomial: Nominal | $\ln(\mu_i)/\ln(\mu_K)$, $i$ ' 1, 2, ...,$K$&1. | $\mu\&\mu\mu$) |
| Multinomial: Ordinal | $\ln\left(\sum_{r=1}^{i}\mu_i\right)\Big/\ln\left(1\&\sum_{r=1}^{i}\mu_i\right)$, $i$ ' 1, 2, ...,$K$&1. | $\mu\&\mu\mu$) |
| Survival | Not applicable | |

| Table 2. Available models with robust variance estimation | | |
|---|---|---|
| Model | GENMOD | SUDAAN |
| Linear | Yes | Yes |
| Logit | Yes | Yes |
| Probit | Yes | No |
| Poisson | Yes | Coming soon |
| Gamma | Yes | No |
| Multinomial: Nominal | No | Yes |
| Multinomial: Ordinal | No | Yes |
| Survival | No | Yes |

Table 2 presents availability of some of the models with robust variance estimation.

## Variance estimation

We have discussed five methods of variance estimation, which are: Model based, Binder, Zeger, Jackknife, and BRR. SUDAAN offers all of these, but GENMOD has only two, model based and Zeger. SUDAAN plans to implement Jackknife and BRR in MULTILOG and SURVIVAL.

| Table 3. Available Variance estimation methods | | |
|---|---|---|
| Method | GENMOD | SUDAAN |
| Model Based | Yes | Yes |
| BINDER | No | Yes |
| Zeger | Yes | Yes |
| Jackknife | No | Yes |
| BRR | No | Yes |

## Correlation structures

In most data analysis problems the estimates $\hat{\beta}$ change very little, when independent working correlation is replaced by exchangeable working correlation. The changes are still smaller when more complex forms of working correlations are used instead of exchangeable working correlation. Furthermore, the robust variance estimator is valid irrespective of the working correlation that you specify.

GENMOD permits you specify one of five possible working correlation structures. However, in SUDAAN you are limited to one of the two choices: Independent or exchangeable working correlation. Table 4 provide a quick summary of available working correlation structures.

| Table 4. Available working correlation structures | | |
|---|---|---|
| Correlation Structure | GENMOD | SUDAAN |
| Fixed | Yes | Yes |
| Independent | Yes | Yes |
| exchangeable | Yes | Yes |
| m-dependent | Yes | No |
| unstructured | Yes | No |
| Autoregressive | Yes | No |

## Tests of hypotheses

GENMOD uses only the Wald chisquare statistic and significance level (or p-value) based on it. SUDAAN provides several alternative adjusted F-statistics for the tests of hypotheses and reports significance levels based on them. Tests based on F-statistics are relevant when the degrees of freedom associated with the variance estimate is small, when number of clusters is small (less than 30). F-statistic provides correct value for the significance level or (p-value).

In SAS release 6.12, the tests of hypothesis for betas are based on correct estimates of variances, but tests for effects, interactions, and contrasts, use the variances based on independent working correlation.

## Estimation of correlation

The estimation of the correlation is slightly different in the two programs. Both GENMOD and SUDAAN estimate the correlation within a cluster I as

$$?_i \text{ ' } \frac{1}{n_i(n_i\&1)}\sum_{j\neq k} e_{ij}e_{ik}.$$

However the programs average these differently to arrive at the overall estimate for the exchangeable correlation. GENMOD takes a simple average of the within cluster correlations, whereas SUDAAN takes weighted average, with weights proportional to $(n_i\&1)$:

$$?_G \text{ ' } \frac{1}{H}\sum_{i=1}^{H} ?_i,$$

$$?_S \; ' \; \frac{1}{n_{\%}\&H} \; \mathbf{j}_{i\,1}^{H} \; (n_h\&1)?_i,$$

where $n_{\%} \; ' \; \mathbf{j}_{i\,1}^{H} \; n_i$. If the number of observations per cluster are the same, or the correlation is the same from cluster to cluster, there is no difference in these methods. When cluster sizes are different, SUDAAN uses greater weight for correlations from larger clusters, because the correlations from the larger clusters have smaller variances and for averaging an estimate, the weights should be inversely proportional to the variances.

GENMOD requires that the robust estimation of variance be carried out at the same level at which the working correlation is estimated, SUDAAN permits you to specify a level different than the one for which correlation is assumed. To illustrate with an example consider the psychological experiment on animals, where repeated measures of a response to given stimulus are made on the same animal. Now our observations have potentially two different levels of correlations: between the observations on the same animal, and between different animals from the same litter. In SUDAAN, you can request the exchangeable correlations at the animal level, and robust estimation at the litter level. The working correlation is such that correlations between repeated measures within animals is exchangeable and correlation between animals within the litter is independent. The robust variance estimation is valid for any arbitrary correlation structure between and within animals.

One of the other differences, is inherent in the algorithm for the iterative solution of the estimating equations. All programs use the modified Newton Raphson algorithm, but its implementation in the computer code differs. The differences are not significant, when there is a single function, such as maximum log likelihood function. However, in solving estimating equations, there is not a single function to be maximized, but there is only a set equations to be solved. Each package uses its own heuristic criterion to decide if the current iteration has produced an improvement over the previous one, and to decide if the solution has converged. You should make sure that the solution provided by a program has properly converged for your data.

## Conclusion

GENMOD and SUDAAN have strengths in some aspects. GENMOD permits wide variety of link and variance functions, within the GLM class. SUDAAN has fewer functions but has procedures for multinomial logistic and survival models. GENMOD permits several choices for working correlation structure, whereas SUDAAN allows only independent and exchangeable working correlations. Robust Variance estimation in GENMOD is based only on Zeger's method. SUDAAN implements, four different methods for robust estimation of variances: Binder, Zeger, Jackknife. and BRR.

Depending on your choice of the model and preference for variance estimation, you should choose the most appropriate tool for your data analysis. In the appendix, we present some examples, biased in favor of SUDAAN. For examples on PROC GENMOD, you may refer to SAS user's manual.

## References


Bieler, G.S. and Williams, R.L. (1997) Analyzing repeated measures and cluster-correlated data using SUDAAN Release 7.5. Research Triangle Institute, Research Triangle Park, NC 27709. (Presented at the joint statistical meetings continuing education workshop.)

Binder, D.A. (1983). "On the Variances of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review, 51, 279-292*.

Binder, D.A. (1992). "Fitting Cox's Proportional Hazards Models from Survey Data" Biometrika 79 139-147.

Folsom, R.E. (1974). "National Assessment Approach to Sampling Error Estimation, Sampling Error Monograph," Prepared for National Assessment of Educational Progress, Denver, CO.

Kish, L., and M. R. Frankel (1974). "Inference from Complex Surveys," Journal of Royal Statistical Society B, 36, 1-37.

Krewski, D. and Rao, J.N.K (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9, 25-45.

McCullagh P.(1989), "Qausi-likelihood functions," Annals of statistics 11, 59-67

McCullagh P. and Nelder, J.A. (1989), Generalized linear models. Chapman and Hall, London, England.

Nelder, J.A. and Wedderburn, R.W.M. (1972), "Generalized linear models," Journal of royal statistical society A, 135, 370-384

Rao, J.N.K. and Scott, A.J. (1987), "On simple adjustment to chi-squared tests with sample survey data," Anaals of statistics, 15, 385-397.

Rao, JNK, Wu, CFJ, and Yue, K (1992). "Some Recent Work on Resampling Methods for Complex Surveys." *Survey Methodology*, `8, . 209-217.

Shah, B.V., M.M. Holt, and R.E. Folsom (1977), "Inference about Regression Models from Survey Data," Bulletin of the International Statistical Institute XLVII 3, 43-57.

Shah, BV, Barnwell, BG, and Bieler, GS (1997). ***SUDAAN User's Manual, Release 7.5.*** Research Triangle Park, NC: Research Triangle Institute.

Thomas, D.R. and Rao, J.N.K. {1987), "Small sample comparision of level ans power for simple goodness-0f-fit statistics under cluster sampling." Journal of American Statistical Association 82, 630-636.

Zeger, S. and Liang, K.Y. (1986), "Longitudinal data analysis for discrete and continuous outcomes," Biometric 42, 121-130.


## Acknowledgment


I am grateful to Dr. Beth Barnwell, Ms. Gayle Bieler, and Dr. Rick Williams, for their comments and suggestions.


## Appendix
## Examples of SUDAAN procedures.

We present here two examples for multinomial logistic and survival models with SUDAAN procedures.

**Application of SUDAAN to a Cross-Over Clinical Trial With Ordinal Outcomes: Evaluation of a New Inhaler Device**

Qualitative responses in a cross-over clinical trial are often ordinal. Such responses might be, for example, relief, slight relief, or no relief in studies of painkiller effectiveness. Due to the nature of cross-over studies, repeated measurements on the same subject are likely to be correlated. The intra-subject correlation must be taken into account in order to make valid inferences about the treatment effect.

Data for this example are from a two-treatment two-period crossover study conducted by 3M Health Care Ltd (Ezzet and Whitehead, 1991) to compare the suitability of two inhalation devices (A and B) in patients who are currently using a standard inhaler device delivering salbutamol. The first sequence of patients were randomized to Device A for one week (period 1) followed by Device B for another week (period 2). The second sequence of patients received the treatments in the opposite

order (Device B in period 1, Device A in period 2). Patients gave their assessment on clarity of leaflet instructions accompanying the devices, recorded on an ordinal scale of: 1 = *easy*, 2 = *clear only after re-reading*, 3 = *not very clear*, and 4 = *confusing*.

For the analysis of the data on the clarity of the leaflet data, SUDAAN MULTILOG procedure is most appropriate. Actual statements are presnted in Table 5. For detailed explanation and results, please download Bieler and Williams (1997) for Research Triangle institute's web page.

Table 5.
SUDAAN statements for MULTILOG for evaluation of inhaler device

```
PROC MULTILOG DATA="C:\\TERA\\GEEORD\\CROSS"
FILETYPE=SAS SEMETHOD=ZEGER  R=INDEPENDENT;
NEST _ONE_ PERSON;
WEIGHT _ONE_;
SUBGROUP CLARITY TREAT PERIOD;
LEVELS  4     2    2;

MODEL CLARITY = TREAT PERIOD / CUMLOGIT;
TEST WALDCHI;

TITLE "PROPORTIONAL ODDS MODEL FOR INHALER
DEVICE CROSS-OVER STUDY"
  "Ezzett and Whitehead, 1991";
```

**Using SUDAAN to Analyze a Cross-Over Clinical Trial with Multivariate Failure Time Data:Evaluation of a Coronary Heart Disease Drug on Repeated Exercise Times to Angina Pectoris**

This example demonstrates SUDAAN's correlated data techniques in the context of a clinical trial. The data for this example represent repeated exercise times (in seconds) to angina pectoris in patients with coronary heart disease. We analyzed the data reported by Crouchley and Pickles (1993), in which 21 subjects were each tested four times on one day and a further four times two days later. On each day exercise time measurements were taken just before and at 1 hour, 3 hours, and 5 hours following drug administration. On one day the drug was an active treatment (an oral dose of isosorbide dinitrate) and on the other placebo. Although undertaken as a double-blind randomized cross-over design, the published data do not indicate the order of treatment, preventing any testing for carry-over effects.

The Cox proportional hazards model was used to evaluate the regression effect of treatment (or test day), after adjusting for several covariates: time since drug administration (4-level factor), and indicators for previous myocardial infarction (MI), previous coronary artery bypass surgery (CAB), and previous propranolol treatment (PP). Note that treatment day and time since drug administration are *within-cluster* covariates, while MI, PP, and CAB represent *cluster-level* covariates. For comparison, we include results based on assuming complete independence among the 8 failure times per subject.

The SUDAAN program contains code to fit the Cox proportional hazards model to the observed event times. The default sample design option *DESIGN=WR* (notation for "with-replacement sampling") invokes the robust variance estimator that is appropriate for the study. The *NEST* statement in SUDAAN indicates that the patient (PATIENT) represents the cluster or primary sampling unit, with the keyword *_ONE_* indicating there is a single design stratum. Additional sources of intracluster correlation, such as time within each study day, need not be specified. The requested test statistics *WALDCHI* and *SATADJCHI* refer to the Wald chi-square test and the Satterthwaite-adjusted chi-square test (Rao and Scott, 1987), respectively. The latter test is a modification of the Wald statistic and has been shown to have superior operating characteristics for multiple-degree-of-freedom hypotheses in small samples (Thomas and Rao, 1987). The SUDAAN statements for the PROC SURVIVAL are in Table 6.

Table 6.
SUDAAN statements for Evaluation of a Coronary Heart Disease Drug on Repeated Exercise Times to Angina Pectoris

```
PROC SURVIVAL DATA="EXERCISE" FILETYPE=SAS;
NEST _ONE_ PATIENT;
WEIGHT _ONE_;
SUBGROUP HRS SUDTRT;
LEVELS  4  2;

EVENT COMPLETE;
MODEL EXTIME = SUDTRT HRS MI CAB PP;
EFFECTS MI CAB PP / NAME = "Combined Effect:
MI,CAB,PP";
TEST WALDCHI SATADJCHI;

TITLE "EXERCISE TIME TO ANGINA PECTORIS
(SECONDS):  PLACEBO VS. TREATMENT"
"PROPORTIONAL HAZARDS REGRESSION USING
ROBUST VARIANCE ESTIMATOR:"
 "Main Effects Model";
FOOTNOTE "Source:  Crouchley and Pickles (1993, Biometrics
49, 1067-1076)";
```



Babubhai V. Shah
Chief Scientist, Statistics Research Division
Research Triangle Institute
PO Box 12194
Research Triangle Park NC 27709
Phone: 919-541-6879
Fax: 919-541-7431
E-mail: shah@rti.org

# Software for GEE: PROC GENMOD and SUDAAN

Babubhai V. Shah, Research Triangle Institute, Research Triangle Park, NC

Relevant Key words:

- Generalized Estimating Equations
- GEE
- GENMOD
- SUDAAN
- Correlard data
- Clustered data
- Generalized linear models
- Jackknife
- Balanced repeated replications (BRR)