Paper 275-25

# Multiple Comparisons with a Control in GEE Models Using the SAS System

Jean G. Orelien, Analytical Sciences, Inc., Durham, North Carolina
Jun Zhai, Analytical Sciences, Inc., Durham, North Carolina
Richard Morris, Analytical Sciences, Inc., Durham, North Carolina

## ABSTRACT

In toxicology, it is often of interest to perform a Dunnett's test to determine whether a pesticide or toxic substance is different at various levels from a control. The problem is complicated by the fact that often a traditional analysis of variance (ANOVA) model cannot be used because of a complex variance structure. For example, it might be of interest to determine whether rats receiving different doses of a toxic substance are statistically different with respect to certain outcome measures from a control. Because of concerns that animals from the same litter might be correlated, generalized estimating equations (GEE) methods are often used in modeling these types of data. In SAS, PROC GENMOD allows GEE fitting but does not provide p-values for multiple comparisons with a control. In this paper, we discuss how the algorithm proposed by Hsu (1992) for performing the Dunnett's test and implemented in PROC GLM in the LSMEANS statement can be extended to GEE models. A SAS macro is provided for computing p-values and Confidence Intervals (CI) similar to those obtained in PROC GLM for the two-sided Hsu-Dunnett test.

## I INTRODUCTION

The statistical literature is abundant on the use of GEE methods to model correlated data. In teratology experiments, often a xenobiotic agent is administered to dams and measurements on various endpoints are taken on the pups. Because of concerns that observation on pups from the same dam might be correlated, statistical methods such as GEE are used rather than a traditional general linear model (GLM). It is often of interest to researchers in these experiments to test simultaneously whether the effects of a compound at different dose levels are significantly different from a control. For a GLM or mixed model, tests for multiple comparison with a control are readily available and are implemented in the SAS System. However for GEE models, tests for multiple comparisons with a control are not yet available.

In this paper we propose an algorithm to obtain p-values for a 2 sided multiple comparison test with a control in GEE models similar to Dunnett's test in GLM. First, in section II we describe how the Dunnett's test is computed in SAS in the GLM case. In section III, we show how the algorithm used in the LSMEANS statement of PROC GLM to compute p-values for the Dunnett's test can be extended to GEE models. In section IV, we give an example showing how the method suggested in section III can be implemented in SAS. A SAS macro described in section V is provided in the appendix for computing p-values and CI similar to those obtained in PROC GLM for multiple comparisons with a control for a one factor treatment. The macro can easily be modified if there are more than one treatment factor. Although the algorithm described in this paper is for 2 sided Dunnett test, it can be easily applied to the one sided case. The methodological approach used in section III could also be implemented in other statistical software and particularly in any matrix language software.

## II Dunnett's Test in PROC GLM

Consider an experiment with $k$ treatments and one control. The experimenter wishes to test the differences between the control and each of the treatment with a family wise error rate of $\alpha$:

$$H_o: u_i = u_0, \quad H_a: u_i \neq u_0 \text{ for } i = 1, 2, \ldots, k. \qquad (1)$$

Let $\sigma^2_i$ be the variance for $\bar{x}_i - \bar{x}_0$ and $u_i$, $u_0$ are respectively the $i^{th}$ treatment effect and the effect of the control. Define $t_i$ as follows:

$$t_i = \frac{|\bar{x}_i - \bar{x}_0|}{\hat{\sigma}_i} \qquad (2)$$

Under the null hypothesis, $t_i$ has a t distribution. Suppose that d is such that:

$$Prob(t_1 < d,\ t_2 < d, \ldots,\ t_p < d) = 1 - \alpha \qquad (3)$$

(3) is equivalent to:

$$Prob(z_1 < d * \hat{\sigma}_1,\ z_2 < d * \hat{\sigma}_2,\ \ldots,\ z_k < d * \hat{\sigma}_k) = 1 - \alpha \qquad (4)$$

In the GLM model, $\hat{\sigma}_i = \sqrt{s^2 * (\frac{1}{N_i} + \frac{1}{N_0})}$ where $s^2$ is the pooled variance and $N_i$, $N_0$ are respectively the sample size for the $i^{th}$ treatment and the sample size for the control.

$\begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_k \end{pmatrix}$ has a multivariate normal distribution with mean **0**. The variance for any $z_i$ is $\sigma^2{}_i$ and the correlation between $z_i$ and $z_j$ is $\rho_{ij}$. Now assume that $\rho_{ij}=\lambda_i*\lambda_j$. The correlation matrix for Z is then given by:

$$R=\begin{bmatrix} 1 & \lambda_1*\lambda_2 & \dots & \dots & \lambda_1*\lambda_k \\ \lambda_2*\lambda_1 & 1 & \dots & \dots & \lambda_2*\lambda_k \\ \dots & \dots & \dots & \dots & \dots \\ \lambda_k*\lambda_1 & \dots & \dots & \dots & 1 \end{bmatrix}= \qquad (5)$$

$$\begin{bmatrix} 1-\lambda^2{}_1 & 0 & 0 & \dots & 0 \\ 0 & 1-\lambda^2{}_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1-\lambda^2{}_k \end{bmatrix} + \begin{bmatrix} \lambda^2{}_1 & \lambda_1*\lambda_2 & \dots & \dots & \lambda_1*\lambda_k \\ \lambda_2*\lambda_1 & \lambda^2{}_2 & \dots & \dots & \lambda_2*\lambda_k \\ \dots & \dots & \dots & \dots & \dots \\ \lambda_k*\lambda_1 & \dots & \dots & \dots & \lambda^2{}_k \end{bmatrix} \qquad (6)$$

It can be shown that

$$Z'=\left(\frac{\bar{x}_1-\bar{x}_0-(u_1-u_0)}{\sigma_1}, \frac{\bar{x}_2-\bar{x}_0-(u_2-u_0)}{\sigma_2}, \dots, \frac{\bar{x}_k-\bar{x}_0-(u_k-u_0)}{\sigma_k}\right)$$

has the same distribution as:

$$\left(\sqrt{1-\lambda^2{}_1}Z_1+\lambda_1*Z_0, \dots, \sqrt{1-\lambda^2{}_k}Z_k+\lambda_k*Z_0\right)$$

where $Z_0, Z_1, \dots, Z_k$ are iid standard normal random variables. To see the above relationship notice that under $H_o, (u_i - u_0)=0$ and so

$$Z'=B*\begin{pmatrix} \bar{x}_1-\bar{x}_0 \\ \bar{x}_2-\bar{x}_0 \\ \dots \\ \bar{x}_k-\bar{x}_0 \end{pmatrix}=B*\begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_k \end{pmatrix} \text{ Where } B=\begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \frac{1}{\sigma_k} \end{bmatrix}$$

According Hsu and Nelson (1998), by conditioning on the pooled standard deviation $\hat{\sigma}$ and $Z_0$, $d$ can be written as a solution to:

$$\int_0^\infty\int_{-\infty}^{\infty}\prod_{i=1}^{k}\left[\Phi\left(\frac{\lambda_i y+du}{\sqrt{1-\lambda^2{}_i}}\right)-\Phi\left(\frac{\lambda_i y-du}{\sqrt{1-\lambda^2{}_i}}\right)\right]d\Phi(y)\Upsilon(u)du=1-\alpha \qquad (7)$$

where $\Phi$ is the standard normal distribution and $\Upsilon$ is the density of $\frac{\hat{\sigma}}{\sigma}$. With infinite degrees of freedom (7) becomes:

$$\int_{-\infty}^{\infty}\prod_{i=1}^{k}\left[\Phi\left|\frac{\lambda_i y+d}{\sqrt{1-\lambda^2{}_i}}\right|-\Phi\left|\frac{\lambda_i y-d}{\sqrt{1-\lambda^2{}_i}}\right|\right]d\Phi(y)=1-\alpha \qquad (8)$$

Thus, to obtain p-values for testing whether any of the level of a treatment is different from the control, we can simply use (7) or (8). For a given level *i*, p-values are obtained by computing 1 minus the left side of the expression in (7) or (8) using the values for the $\lambda_i's$ and substituting $d_i=(|(\bar{x}_i - \bar{x}_0)|)/\hat{\sigma}_i$ for the value of *d*. That is for a given level *i*, p-values would be computed as:

$$1 - \int_0^\infty\int_{-\infty}^{\infty}\prod_{i=1}^{k}\left[\Phi\left(\frac{\lambda_i y+(d_i)u}{\sqrt{1-\lambda^2{}_i}}\right)-\Phi\left(\frac{\lambda_i y-(d_i)u}{\sqrt{1-\lambda^2{}_i}}\right)\right]d\Phi(y)\Upsilon(u)du \qquad (9)$$

or

$$1 - \int_{-\infty}^{\infty}\prod_{i=1}^{k}\left[\Phi\left(\frac{\lambda_i y+d_i}{\sqrt{1-\lambda^2{}_i}}\right)-\Phi\left(\frac{\lambda_i y-d_i}{\sqrt{1-\lambda^2{}_i}}\right)\right]d\Phi(y) \qquad (10)$$

where (9) is for finite degrees of freedom and (10) is for the case when we are assuming infinite degrees of freedom.

For the case *k*=3, that is when there are 3 treatment levels and a control, the values for the $\lambda_i's$ can be determined exactly by solving:

$$\rho_{12}=\lambda_1*\lambda_2$$
$$\rho_{13}=\lambda_1*\lambda_3$$
$$\rho_{23}=\lambda_2*\lambda_3$$

The same is true for the case *k*=2. When *k* is greater than 3, factor analysis can be used to estimate the values of the $\lambda_i s$ (Hsu, 1992). These values for the $\lambda_i s$ are not exact and so the p-values are only an approximation. In SAS, factor analysis with the iterated principal factor method is used to obtain the values of the $\lambda_i s$.

## III Extending the Algorithm Used to compute the Dunnett's Test in PROC GLM to GEE models

Now assume that we have *k* treatments and a control but that the variance structure is more complex than in the traditional GLM case, i.e., we have a model $f(Y)=XB+\varepsilon$ where $f(.)$ is the link function and the structure of the error term is more complex than the traditional case. With teratology data, we assume that the total variance is made up of within litter variances and between litter variances. Assume as in section II that we have *k* treatments and a control and that we desire to test: $H_o$: $u_i=u_0$, $H_a$: $u_i\neq u_0$ for *i*=1, 2, ..., *k*.

Let $l_i'$ denotes the vector containing the coefficients corresponding to a univariate test of the control vs. the $i^{th}$ treatment. For example with 3 treatments and a control, $l_1'=(0, 1, -1, 0)$. Now let L be a matrix which rows are made up of the vectors $l_i'$. For example, with 3 treatments and a control with the parameterization used in SAS, we have:

$$L' = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The variance-covariance matrix for the set of contrasts defined by L can be approximated with:

$$V = L' \Sigma L$$

where $\Sigma$ is the empirical variance-covariance matrix for the betas.

For every treatment *i* defines the test statistic to be:

$$d_i = \frac{|(\hat{u}_i - \hat{u}_0)|}{\sqrt{var(u_i - u_0)}} = \frac{|l_i' * \hat{\beta}|}{\sqrt{l_i' \Sigma l_i}} \qquad (11)$$

The p-value for any of the *k* tests for a family wise error rate $\alpha$ is given by:

$$pvalue = 1 - prob(z_1 < d_1, \ ..., \ z_k < d_k) \qquad (12)$$

where the z's have a multivariate normal joint distribution with mean **0** and variance-covariance matrix the correlation matrix of $L' \Sigma L$.

If we assume that $corr(u_i - u_0, \ u_j - u_0) = \rho_{ij} = \lambda_i * \lambda_j$ then we can compute the p-value in (12) by using (10). Similar to what is done in SAS in PROC GLM we can obtain values for the $\lambda_i s$ by using factor analysis.

The SAS function PROBMC(*test, q, prob, df, nparms<parms>*). computes (10). The value for *test* is set to "DUNNETT2", the value for *q* is set to $d_i$, the value for *prob* and the value for *df* is set to missing, the number of parameters is set to *k*, that is the number of treatments (excluding the control) and the parameters are the values of the $\lambda_i s$. The $\lambda_i s$ can be computed using PROC FACTOR with the iterated principal factor method. Quantiles can be obtained using PROBMC by setting *q* to missing and setting *prob* to be the percentile for which a quantile is requested. Using a value of *prob*=0.95 would yield the *q* value to be used so that one can obtain 95% simultaneous confidence intervals for $u_i - u_0$:

$$(\bar{x}_i - \bar{x}_0 - q * s_i, \ \ \bar{x}_i - \bar{x}_0 + q * s_i)$$

To get a sense of how good the approximation of the correlation matrix in terms of the lambdas is, one can look at the residual matrix:

$$\boldsymbol{R} - (\boldsymbol{I} - diag(\boldsymbol{\lambda}\boldsymbol{\lambda}') + \boldsymbol{\lambda}\boldsymbol{\lambda}')$$

(Recall that $\boldsymbol{R}$ is the correlation matrix. $\boldsymbol{\lambda}$ is a vector containing the values of the $\lambda_i s$). The idea behind this is that if we could factor R, we would have $\boldsymbol{R} = (\boldsymbol{I} - diag(\boldsymbol{\lambda}\boldsymbol{\lambda}') + \boldsymbol{\lambda}\boldsymbol{\lambda}')$ and so with an approximation of R, we would expect $\boldsymbol{R} - (\boldsymbol{I} - diag(\boldsymbol{\lambda}\boldsymbol{\lambda}') + \boldsymbol{\lambda}\boldsymbol{\lambda}')$ to be close to **0**.

## IV Example

In this section, we give an example of how p-values can be computed for a GEE model using the algorithm presented in section III. The data used is from a toxicology experiment from the national toxicology program (NTP) to study the effects of methoxychlor on male rats reproductive organs. There are 5 doses group: 0, 0.05, 0.5, 5, 50 and 150. It is assumed that measurements on male rats from the same litter are correlated. PROC GENMOD was used to analyze the data with an exchangeable correlation structure. Epididis weight is the outcome variable and dose of methoxychlor is the independent variable. The raw data is in the appendix as part of the SAS code.

The beta parameters $\hat{\beta}$ is:

$$\hat{\beta} = \begin{bmatrix} 0.2279 \\ -0.0011 \\ 0.0179 \\ -0.0664 \\ 0.0213 \\ -0.0855 \end{bmatrix}$$

The first element of the beta vector corresponds to the intercept, the second element corresponds to dose group=5, the 3rd to dose group=50, the fourth to dose group=0.5 and the fifth to dose group=150. Dose group=0.05 is used as the reference cell.

The empirical covariance matrix of the betas, $\Sigma$ is:

$$\Sigma = \begin{bmatrix} 0.0001489 & -0.000149 & -0.000149 & -0.000149 & -0.000149 & -0.000149 \\ -0.000149 & 0.000431 & 0.0001489 & 0.0001489 & 0.0001489 & 0.0001489 \\ -0.000149 & 0.0001489 & 0.0003065 & 0.0001489 & 0.0001489 & 0.0001489 \\ -0.000149 & 0.0001489 & 0.0001489 & 0.0004722 & 0.0001489 & 0.0001489 \\ -0.000149 & 0.0001489 & 0.0001489 & 0.0001489 & 0.0003147 & 0.0001489 \\ -0.000149 & 0.0001489 & 0.0001489 & 0.0001489 & 0.0001489 & 0.0001489 \end{bmatrix}$$

The first column of $\Sigma$ is the intercept, the other columns are respectively the control, treatment 1, treatment 2, treatment 3 and treatment 4. Treatment 5 is the reference cell.

The $L'$ matrix is given by:

$$L' = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Essentially $L'$ corresponds to the set of contrast that compares the control to each dose level. The variance-covariance matrix for the contrasts is $L'\Sigma L$. The correlation matrix of $L'\Sigma L$ is:

$corr(L'\Sigma L) =$

| | | | | |
|---|---|---|---|---|
| 1 | 0.5467580212 | 0.635644047 | 0.8009646091 | 0.6479847325 |
| 0.5467580212 | 1 | 0.541729509 | 0.6826244444 | 0.5522468945 |
| 0.635644047 | 0.541729509 | 1 | 0.7935981688 | 0.6420252422 |
| 0.8009646091 | 0.6826244444 | 0.7935981688 | 1 | 0.8090054481 |
| 0.6479847325 | 0.5522468945 | 0.6420252422 | 0.8090054481 | 1 |

The $\lambda_i s$ are obtained by inputting the matrix above in PROC FACTOR with the iterated principal option. The test statistic for each dose level can be computed using (11). In table 4-1, we give the values for the test statistic and the p-values. The p-values are obtained by using the SAS function *PROBMC*:

$pvalue = 1 - PROBMC(TEST, d_i, ., 5, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ where test is a constant character string whose value is "DUNNETT2" and $d_i$, the test statistic for the $i^{th}$ treatment is the difference between the $i^{th}$ treatment level and the control divided by the standard error of the difference. The lambda values in this example are $\lambda_1 = 0.80107$, $\lambda_2 = 0.68268$, $\lambda_3 = 0.7937$, $\lambda_4 = 0.99958$ and $\lambda_5 = 0.80911$.

**Table 4-1. P-values for Multiple Comparison With a Control from a GEE Model**

| Dose | Test Stat. | P-value |
|---|---|---|
| 0.05 | 0.0533 | 1 |
| 0.5 | 1.0596 | 0.6492 |
| 5 | 0.9061 | 0.7589 |
| 50 | 2.6538 | 0.0286 |
| 150 | 5.0279 | < 0.001 |

We computed $\boldsymbol{R} - (\boldsymbol{I} - diag(\boldsymbol{\lambda\lambda}') + \boldsymbol{\lambda\lambda}')$ to see how well the correlation matrix could be approximated by $\boldsymbol{I} - diag(\boldsymbol{\lambda\lambda}') + \boldsymbol{\lambda\lambda}'$. The entries in the residual matrix are all closed to 0 suggesting that $\boldsymbol{I} - diag(\boldsymbol{\lambda\lambda}') + \boldsymbol{\lambda\lambda}'$ provides a good approximation of $R$.

$\boldsymbol{R} - (\boldsymbol{I} - diag(\boldsymbol{\lambda\lambda}') + \boldsymbol{\lambda\lambda}') =$

| | | | | |
|---|---|---|---|---|
| 2.220446E-16 | -0.000115198 | -0.000163194 | 0.000232904 | -0.000172087 |
| -0.000115198 | -1.11022E-16 | -0.000111866 | 0.0002327686 | -0.000118917 |
| -0.000163194 | -0.000111866 | 2.220446E-16 | 0.0002340895 | -0.000167811 |
| 0.000232904 | 0.0002327686 | 0.0002340895 | -2.22045E-16 | 0.0002314902 |
| -0.000172087 | -0.000118917 | -0.000167811 | 0.0002314902 | -2.22045E-16 |

## V Description of the SAS Macro

A SAS macro Dunnett_GEE is provided in the appendix. This macro was written in SAS version 7 and may not work with earlier versions. The only information required as input are the beta vector, the empirical variance-covariance matrix of the betas and the number of treatments excluding the control. The macro assumes that the beta vector and variance-covariance matrix are created using PROC GENMOD. PROC IML is used to compute the test statistics and the correlation matrix for the parameters made up of the differences between each treatment and the control. PROC FACTOR is then used to decompose the correlation matrix to obtain the lambdas. In a data step, the lambdas from PROC FACTOR and the test statistics computed in IML are inputted into PROBMC to obtain p-values and the critical value necessary to obtain 95% simultaneous confidence intervals.

## VI Conclusion

GEE models are often used to analyze data from teratology experiments in order to account for the complex variance structure. In these experiments, one question of interest to researchers is whether several doses of a compound are significantly different from a control. A test similar to Dunnett's in the traditional GLM is not available for GEE models. In this paper, we extend the algorithm proposed by Hsu (1992) for multiple comparisons with a control to GEE models and provide a SAS macro to perform the computations for a one way design.

## References:


Dunnett, C. W. (1955). "A Multiple Comparison Procedure for comparing Several Treatments with a control.", JASA, 50, 1096-1121.

Gupta, S. S. (1963). "Probability Integrals of Multivariate Normal and Multivariate t" Annals of Mathematical Statistics, 34, 792-828.

Hsu J. C. (1992). "The Factor Analytic Approach to Simultaneous Inference in the General Linear Models". Journal of Computational and Graphical Statistics, 1, 151-168

Hsu J. C., Nelson B. (1998). "Multiple Comparisons in the General Linear Model", Journal of Computational and Graphical Statistics, 7, 23-41.

Royen, T. (1987) "An approximation for Multivariate Normal Probabilities of Rectangular Regions", Statistics, 18, 389-400.

Zbynek, S. (1967). "On Multivariate Normal Probabilities of rectangles: Their Dependence on Correlations", The Annals of Mathematical Statistics, 39, 1425-1434.

SAS Institute. *SAS/STAT Software: Changes and Enhancements through Release 6.12*, Cary, NC: SAS Institute, Inc.

## Acknowledgments

We thank Dr. Robert Chapin and Martha Harris of NIEHS for allowing us the use of data from NTP experiments.

## Contact Information

Your questions and comments are welcome. Please contact:

Jean G. Orelien
Analytical Sciences, Inc.
2605 Meridian Pkwy.
Durham, NC 27713
Work Phone: (919)544-8500 (ext. 112)
Fax: (919)544-7307
Email: jorelien@asciences.com
Web: http://www.asciences.com

## Appendix. A Macro to Compute p-values for Multiple Comparison With a Control in GEE Models

```
/*************************************

THIS PROGRAM COMPUTES P-valueS FOR THE DUNNETT_HSU TEST AND 95% CI FOR GEE MODELS.
THE INPUTS FOR THIS PROGRAM ARE THE BETA VECTOR, THE VARIANCE-COVARIANCE MATRIX FOR THE BETAS
AND THE NUMBER OF TREATMENTS EXCLUDING THE CONTROL
THIS PROGRAM ASSUMES THAT THE FIRST BETA PARAMETER CORRESPONDS TO THE CONTROL

*************************************/
/*
THE PARAMETERS THAT HAVE TO BE INPUT INTO THE MACRO ARE THE FOLLOWING:
BVECTOR: THE VECTOR OF THE BETAS
COVAR: THE EMPIRICAL VARIANCE-COVARIANCE OF THE BETAS
K: THE NUMBER OF TREATMENT LEVELS EXCLUDING THE CONTROL
*/
%macro dunnett (bvector, covar, k);

/* KEEP ONLY THE VARIABLES THAT HAVE THE COEFFICIENTS FOR THE VARIANCE-COVARIANCE MATRIX */

%let numlev =%eval(&k+1);

Data covbeta;
Set &covar;
Drop rowname;

Data parmname;
Set &bvector;
If level1='0'  then delete;
If parameter='intercept' then delete;
If parameter='scale' then  delete;
Numlevel=%eval(&k);
K=numlevel;
Treatmentname=parameter||' '||level1;
Keep treatmentname k ;

Data parmname;
Set parmname;
Order=_n_;
/* THE MAIN GOAL OF THIS IML CODE IS TO COMPUTE THE CORRELATION MATRIX FOR THE LSMEANS */

Proc iml;
Use covbeta;
Read all into covb;
Use &bvector;
Read all var {estimate} into b;

L2={0 1 -1, 0 1 0};
L3={0 1 -1 0, 0 1 0 -1 , 0 1 0 0 };
L4={0 1 -1 0 0 ,  0 1 0 -1 0 , 0 1 0 0 -1 , 0 1 0 0 0 };
L5={0 1 -1 0 0 0 , 0 1 0 -1 0 0 , 0 1 0 0 -1 0 , 0 1 0 0 0 -1 , 0 1 0 0 0 0};
L6={0 1 -1 0 0 0 0 , 0 1 0 -1 0 0 0 , 0 1 0 0 -1 0 0 , 0 1 0 0 0 -1 0 ,
0 1 0 0 0 0 -1, 0 1 0 0 0 0 0 };

/* COMPUTE LS MEANS */

beta=b[1:&numlev,];
Est=l&k*b[1:&numlev,];

/* FIX THE VAR-COVARIANCE MATRIX TO ACCOUNT FOR THE PRESENCE OF THE INTERCEPT AND THE SCALE PARAMETERS IN THE BETA
VECTOR */

Cov=covb[1: &numlev,1: &numlev];
Cov_diff=l&k*cov*t(l&k);
Cov_diff=l&k*cov*t(l&k);
Corr_diff=(diag((1/sqrt(cov_diff))))*cov_diff*(diag((1/sqrt(cov_diff))));
stderror=sqrt(vecdiag(cov_diff));
q=(est/stderror)||est||stderror;

print cov est beta;

Create corr_lsmns from corr_diff;
Append from corr_diff;

Create lsmeans from q;
Append from q;

Create covdiff from cov_diff;
Append from cov_diff;
```

```
Quit;

Proc print data=corr_lsmns;
Title5 "correlation matrix for the ls-means";

Data lsmeans;
Set lsmeans;
Order=_n_;
Test_stat=abs(col1);
Lsmeans=col2;
Stderror=col3;
Drop col1-col3;

/* SET UP CORRELATION MATRIX TO BE USED BY PROC CORR */

Data corr_lsmns(type=corr);
Set corr_lsmns;
_type_='corr';
_name_='col'||left(put(_n_, z1.));

/* USE PROC FACTOR TO DECOMPOSE MATRIX */

Ods select factorpattern;
Ods output factorpattern=lambda;

Proc factor data=corr_lsmns method=prinit;

Proc transpose data=lambda out=lambda(keep=l1-l&k) prefix=l;
Var factor1;

Data lambda;
Set lambda;
Do i=1 to &k;
Order=i;
Output;
End;
drop i;


Proc sort data=lambda; by order;

Data probmc;
Merge lambda(in=a) lsmeans(in=b) parmname(in=c);
By order;
If a and b and c;
String='dunnett2';
If k=2 then do;
P-value=1-probmc(string,test_stat,.,.,2,l1,l2);
Q95=probmc(string,.,0.95,.,2,l1,l2);
End;
If k=3 then do;
P-value=1-probmc(string,test_stat,.,.,3,l1,l2,l3);
Q95=probmc(string,.,0.95,.,3,l1,l2,l3);
End;
If k=4 then do;
P-value=1-probmc(string,test_stat,.,.,4,l1,l2,l3,l4);
Q95=probmc(string,.,0.95,.,4,l1,l2,l3,l4);
End;
If k=5 then do;
P-value=1-probmc(string,test_stat,.,.,5,l1,l2,l3,l4,l5);
Q95=probmc(string,.,0.95,.,5,l1,l2,l3,l4,l5);
End;
If k=6 then do;
P-value=1-probmc(string,test_stat,.,.,6,l1,l2,l3,l4,l5,l6);
Q95=probmc(string,.,0.95,.,6,l1,l2,l3,l4,l5,l6);
End;

Lci=lsmeans-q95*stderror;
Uci=lsmeans+q95*stderror;
Drop l&numlev-l9;

Proc print data=probmc;
Title5 "Dunnett-Hsu p-values and 95% CI";

%mend dunnett;

/*  Run macro with Example of Epididis Weight  */

data epidwt;
input @@
DamId PupId Dose EpididisWt;
cards;
165 1  0.00 0.2876 165 2  0.00 0.2332
167 1  0.00 0.219 167 3  0.00 0.1622
172 1  0.00 0.1846 178 1  0.00 0.2312
178 4  0.00 0.2701 168 1  0.05 0.2626
168 2  0.05 0.2068 169 1  0.05 0.2033
169 4  0.05 0.1936 182 1  0.05 0.1925
182 4  0.05 0.237 187 1  0.05 0.2784
187 3  0.05 0.2494 164 1  0.50 0.291
164 2  0.50 0.2415 164 3  0.50 0.23
164 4  0.50 0.2874 185 2  0.50 0.2262
185 4  0.50 0.2195 166 2  5.00 0.2986
175 1  5.00 0.2027 175 4  5.00 0.2246
181 1  5.00 0.2996 181 3  5.00 0.2132
186 2  5.00 0.2439 186 4  5.00 0.2383
176 2 50.00 0.1838 176 3 50.00 0.226
183 2 50.00 0.1525 183 5 50.00 0.1138
188 1 50.00 0.1932 188 4 50.00 0.1
171 1 150.0 0.1518 171 2 150.0 0.133
;
proc print data=epidwt;
title5 "Epididis Wt. from Metachlor Prostate Evaluation Study";

run;

ods select classlevels geercov geercorr parameterestimates;
ods output geerCov=covB;
ods output ParameterEstimates=beta;

proc genmod data=epidwt;
class damid pupid dose ;
model epididisWt=dose /type3 link=identity  covb  ;
repeated Subject=Damid/type=exch maxiter=25000 ecovb ecorrb;
title5 "Using GENMOD to output betas and variance-covariance matrix of
the betas";
run;

%dunnett(beta, covb, 5);

run;
```