

Repeated-measures analysis and PROC MIXED

Shanna Smith & Tor Neilands

University of Texas at Austin

Abstract

Seasoned users of SAS PROC GLM enjoy the use of a powerful tool for repeated measures analysis of variance; yet SAS Institute's more modern MIXED procedure offers several advantages to the data analyst in comparison to the GLM procedure. First, PROC MIXED allows all available data points to be utilized by the investigator, resulting in improved statistical inference and greater power to detect statistical differences for within-subjects effects. Second, PROC MIXED allows users to specify their own variance-covariance matrices, thereby sidestepping restrictive PROC GLM sphericity assumptions. In addition, contrast coding, which enables comparisons across within- and between-subjects effects, is often complex and sometimes impossible in the GLM procedure. The MIXED procedure allows the data analyst to specify all contrasts in a more straightforward manner. To highlight these concepts, an illustrative example from the first author's research is presented, including a brief tutorial on properly structuring data for repeated measures analysis in PROC MIXED.

Drawing Superior Inferences in Longitudinal Analysis:

An Illustration of Mixed Models Analysis

Longitudinal data is often analyzed through a repeated-measures analysis of variance, a specific implementation of a general class of statistical models known as the general linear model (GLM; Searle, 1987). The GLM is well-known, flexible, and conveniently available in popular statistical software programs such as SAS (SAS Institute, 1990); however, the GLM

makes two restrictive assumptions about longitudinal data. As these assumptions are often violated, many researchers have turned to the mixed models approach, which does not require these assumptions.

The first assumption of the GLM, sphericity, posits that the repeated measures of a within-subjects effect should have equal variances and constant covariances (Howell, 1987; Littell, Milliken, Stroup, and Wolfinger, 1996). When both between-subjects and within-subjects effects are present, the variance-covariance matrices of the dependent variables must be equal across the groups that comprise the between-subjects effects. Violations of sphericity can adversely impact the hypothesis tests for within-subjects and between-by-within-subjects effects (Stevens, 1996). When data violate the sphericity assumption, SAS PROC GLM offers three hypothesis tests for each within-subject effect: A test based upon the multivariate ANOVA model (Harris, 1985), and two adjusted univariate tests, the Huynh-Feldt and Greenhouse-Geisser adjusted F statistics (Stevens, 1996). Each test addresses the sphericity assumption in a different way; unfortunately, they do not always agree, leaving investigators uncertain as to which result to trust. In contrast, mixed models avoid the sphericity assumption by allowing the researcher to specify an explicit covariance structure among the repeatedly measured dependent variables included in the analysis. Furthermore, the mixed models approach allows for unequal covariance matrices between grouping factors in an analysis (Littell et al., 1996). As the software performs tests of mean differences using the researcher's chosen covariance structure¹, the problem of choosing which within-subjects hypothesis tests to interpret is removed.

The second assumption of the GLM approach is that any missing data in the study are missing completely at random (MCAR, Little & Rubin, 1987); that is, only a completely random process distinguishes those research participants with missing data from those with complete

data. As the random loss of data should have no effect on the substantive conclusions drawn from the dataset, cases with incomplete or partially missing data are handled through listwise or casewise deletion: if a case has a missing data point at any measurement occasion, the entire case is deleted from the analysis. The ensuing loss of cases may seriously compromise statistical power (Roth, 1994; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997; but see Fairclough, Peterson, & Chang, 1998 for a dissenting opinion), leading to higher standard errors and an increase in non-significant hypothesis test results (Graham et al., 1997).

In addition, the assumption that data are missing completely at random is often violated in longitudinal data sets; for example, in a study assessing the effectiveness of a study-skills intervention, students with poor academic performance may also be less likely to complete the intervention and return for the follow-up assessment. When listwise deletion is performed with such datasets, results of the analysis can only apply to the population of individuals who can and do complete all measurements (Roth, 1994; Arbuckle, 1996; Duncan, Duncan, & Li, 1998; Fairclough, Peterson, & Chang, 1998). That is, we can draw inferences about the effectiveness of the intervention for only those students who completed the intervention – those who likely had better academic skills in the first place.

The mixed models approach to longitudinal analysis is superior in its treatment of missing data. Modern mixed models methods use iterative maximum-likelihood estimation methods rather than the ordinary least-squares approach of GLM (Searle, 1987; Hocking, 1996). The maximum likelihood approach uses all available data points for analysis, and follows the more relaxed assumption that missing data are missing at random (MAR; Little & Rubin, 1987). That is, the pattern of missing data can be predicted by other measured variables in a study

(Arbucke, 1996; Graham et al., 1997, Heitjan, 1997); as in the example of students with poor academic performance who are more likely to miss the follow-up assessment.

To illustrate the superior performance of mixed models in handling the sphericity and missing data problems typical in longitudinal studies, we present an example from a long-term study of married couples and the quality of their relationships.

The PAIR Project Study

This study explored the relationship between long-term marital status and couple-level sexual satisfaction across the first years of marriage (Smith & Huston, 1998). Two frameworks regarding the development of marital distress were examined: the enduring dynamics and disillusionment models (see Huston & Houts, 1998; Huston, Caughlin, Houts, Smith, & George, in press). The enduring dynamics model posits that couples' relationship patterns are well-established when the couple first marries; thus, couples who are initially less satisfied with various aspects of their relationship will be at a higher risk for eventual marital distress or divorce. In contrast, the disillusionment model assumes that all newlywed couples are blinded by romance, and are thus quite satisfied with most aspects of their marriage; over time, however, some couples' illusions will begin to wear away. Couples who begin this process of disenchantment during the early years should be at a higher risk for eventual marital distress or divorce.

Data originated from the Process of Adaptation in Intimate Relationships (PAIR) Project, a four-wave longitudinal study of 168 couples wed in 1981 in Central Pennsylvania. The first wave of data was collected at the second month of marriage; the second and third wave followed in 1982 and 1983. The follow-up study in 1994 and 1995 collected marital status information thirteen years after the couples were wed. Eligible participants were required to (a) speak

English, (b) be in their first marriage, and (c) have no plan to move out of the area within two years. Forty-two percent of contacted respondents agreed to participate in the study.

During the first three data collection phases, husbands and wives were interviewed separately by same-sex interviewers for an average of three hours per session. During the second panel of data collection, spouses used a nine-point scale, ranging from 1 ("very dissatisfied") to 9 ("very satisfied"), on which they retrospectively evaluated their newlywed satisfaction with their sexual relationship (see Vangelisti & Huston, 1994). They then evaluated their current sexual satisfaction with the same procedure. For each score, the interviewer probed for the reason behind the rating, and the reasons for the continuity or change in the rating from the previous phase. During the third panel, in 1983, spouses gave a new current rating and qualitative description of their sexual satisfaction. To simplify the presentation of the analysis in this paper, husbands' and wives' reports were collapsed to create a couple index of sexual satisfaction.

Marital outcome was assessed during the follow-up phase in 1994, using both marital satisfaction and marital stability. For those who remained married, marital satisfaction was assessed with the Marital Opinion Questionnaire (MOQ), adapted from Campbell, Converse, and Rogers (1976). Couples were classified as "Happy" if both spouses evaluated the marriage above neutral; they were classified as "Not Happy" if one or both evaluated the marriage below neutral. Information on the divorced couples was obtained either from court records or from personal information during the interview. Those who divorced between two and seven years of marriage were placed in the "Early Divorced" category; those who divorced between seven and thirteen years were classified as "Later Divorced." The seven year cutoff point between early and later divorces was chosen because the median length of marriage in the United States for those who

eventually divorce is 7.2 years (U.S. Bureau of the Census, 1997) (see Huston, Caughlin, Houts, Smith, & George, in press).

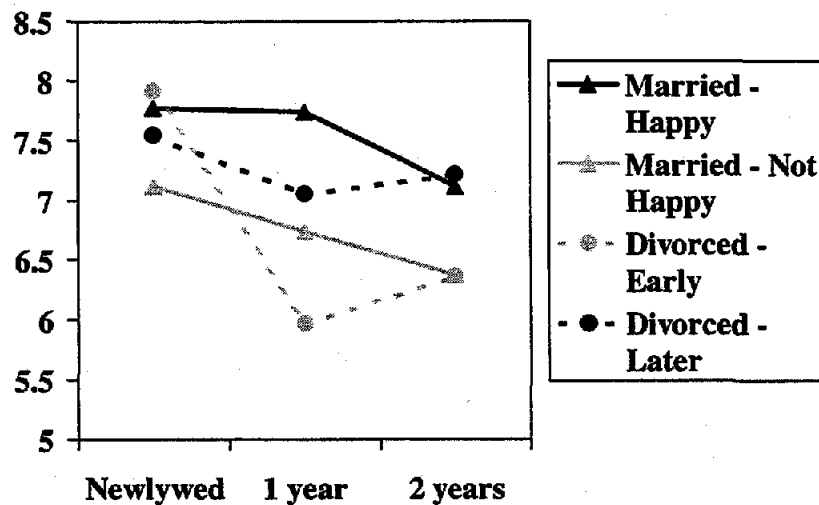
The key issue addressed by the data analysis was whether patterns of early marital sexual satisfaction would differ between the four marital outcome groups (happily married, not happily married, early divorced, and late divorced). The enduring dynamics model would suggest that there would be a main effect for marital status, indicating that the four groups differed in overall sexual satisfaction. The disillusionment model would predict that there would be an interaction between marital status and time, such that those destined for distress or divorce would decline more severely in sexual satisfaction across the early years of marriage than would those who would later be happily married. To address these hypotheses, two analyses were conducted, one using the GLM approach and the other the mixed models approach. Both analyses treated marital status as a four-level between-subjects variable and time as a within-subjects variable.

Results

A total of 156 couples reported their marital status at the follow-up time point and also contributed sexual satisfaction data in at least one panel. Of these cases, 116 individuals had complete data for all three sexual satisfaction measurement occasions. As the GLM employs listwise deletion of cases with incomplete data, the analyses reported first use only the 116 cases with complete data. The analysis was conducted using PROC GLM procedure in SAS 8.1 for Windows (SAS Institute, 1990). Although the GLM analysis omitted only 40 cases, it is highly unlikely that data were missing due to completely random factors.

PROC GLM reported the marital status between-subjects main effect test to be significant, $F(3, 112) = 6.28, p < .001$. However, a statistically significant departure from sphericity was present in the data, Mauchly's $W = .94$, approximate $\chi^2 = 7.11, df = 2, p < .029$,

as well as a significant multisample sphericity violation, Box's $M = 31.89$, $df = 18$, $p < .04$; thus the results that included within-subjects effects needed to be evaluated using one of the adjusted hypothesis. Fortunately, the tests agreed on the significance for each effect: The PROC GLM multivariate test for the within-subjects main effect of phase was statistically significant [Pillai's Trace = .12, $F(2, 111) = 7.51$, $p < .001$; Huynh-Feldt, $F(1.97, 220.15) = 9.40$, $p < .001$; Greenhouse-Geisser, $F(1.88, 210.92) = 9.40$, $p < .003$]. By contrast, the marital status by phase interaction was not significant [Pillai's Trace = .09, $F(6, 224) = 1.68$, $p < .127$; Huynh-Feldt, $F(5.90, 220.15) = 1.67$, $p < .131$; Greenhouse-Geisser, $F(5.65, 210.92) = 1.67$, $p < .135$].



Results indicated that sexual satisfaction decreased with time for all groups. Couples in the Early Divorced and the Married-Not Happy groups appeared to have lower satisfaction with

their sexual relationship across the early years of marriage, as the enduring dynamics model would predict. The means displayed in Figure 1 also hint at an interaction, such that individuals who will divorce early start their marriage with high levels of satisfaction, but wind up with lower levels of satisfaction at phases two and three. Nonetheless, because the interaction effect test was not statistically significant, the investigator would conclude from this analysis that only the main effects were significant findings, and would point to the enduring dynamics model as a complete explanation of the findings.

The analysis described above was replicated using SAS PROC MIXED (Littell et al., 1996; SAS Institute, 1996). Unlike the GLM analysis, the mixed models approach can make use of all available data, so this analysis used a total of 156 cases (see Appendix A for syntax). To deal with the sphericity violation, we chose an unstructured covariance matrix, which allows PROC MIXED to estimate separate values for each variance and covariance in the matrix. Whereas this structure is the most complex and therefore the least parsimonious structure available, all other available covariance structures are a subset of the unstructured covariance matrix; thus, selecting the unstructured matrix is a safe choice if the investigator is uncertain about the properties of the data's covariance structure. This analysis also specified an additional GROUPS option that allowed PROC MIXED to estimate separate covariance matrices for each of the four marital status groups. The authors chose this option because the Box M chi-square test of equal covariance matrices suggested the possibility of a violation of the multisample sphericity assumption.

The mixed models test for the within-subjects main effect of phase was statistically significant, $F(2, 240) = 10.66, p < .001$, as was the marital status between-subjects main effect test, $F(3, 130) = 5.81, p < .001$. In addition, the marital status by phase interaction was

significant, $F(6, 240) = 2.35, p < .032$. As with the GLM analysis, sexual satisfaction decreased with time for all groups, and the unhappily married and early divorced groups appeared to be less satisfied with their sexual relationship across the first years of marriage. The additional statistical power provided by the inclusion of all available cases in the mixed models analysis allows the heretofore non-significant interaction effect to emerge. On the basis of the interaction effect, the researcher could conclude that a synthesis of the emergent distress and disillusionment models are needed to explain the early development of long-term marital distress. Further interpretation of the interaction, however, requires the use of follow-up contrasts. PROC MIXED feature adjusted and unadjusted pairwise comparison options in the LSMEANS statement, as well as user-generated contrasts. In the following section, we illustrate the ease of specifying user-generated contrasts with the PROC MIXED ESTIMATE statement.

Using Contrasts to Interpret Mixed Models Results

Recent work using the same sample of couples (Huston et al., in press) suggests that couples destined for divorce are most likely to feel symptoms of disillusionment across the early years of marriage: their sense of love for their partner declines sharply, and their ambivalence about the relationship increases. In contrast, couples who later become unhappy but remain married are consistently low in love and high in ambivalence. While they do lose some sense of love for one another across the early years, their declines are relatively slight, and parallel to the declines of couples who remain happily married.

If the disillusionment model is more descriptive of couples who divorce, then the divorced couples in our study should experience sharper declines in sexual satisfaction than do the couples who will be headed for a long-term, but unhappy, marriage. While that premise does not appear to hold for the Divorced Later group, who remained relatively stable in their sexual

satisfaction, visual inspection of Figure 1 suggests that the early divorced couples decline more strongly in satisfaction than do either of the married couples.

To determine the significance of the decline in sexual satisfaction across time for the early divorced and unhappily married groups, we appended two single degree of freedom contrasts to the PROC MIXED analysis (see Appendix A). Single degree of freedom contrasts may be carried using either the ESTIMATE statement or the CONTRAST statement in PROC MIXED; the CONTRAST statement is required for multiple degree of freedom contrasts. The first contrast compared the means of the married but not happy group between time 1 and 3; the second compared the means of the early divorced group between time 1 and 3. As expected, the decline in sexual satisfaction for the divorced early group was highly significant, $t = 3.06$; $p = .003$, while that of the married but not happy group was marginal, $t = 1.96$; $p = .051$. Had the investigators chosen to impose a post-hoc adjustment such as the Bonferroni criterion (Hays, 1994), the second contrast would be clearly non-significant whereas the first contrast would remain statistically significant. On the basis of these results, we can conclude that the early divorced group declined more severely in sexual satisfaction across the early years than did those who would remain married but not happy, indicating that the early divorced group showed more symptoms of disillusionment than did the long-term unhappily married group.

Discussion

Previous research has documented the advantages of mixed models for repeated measures longitudinal data analysis from a theoretical standpoint. The current study provides a practical illustration of mixed models' benefits to basic and applied researchers. By using a mixed models analysis with follow-up contrasts rather than the GLM, the investigators draw a substantially

different (and, the authors argue, more correct) conclusion about the data, leading to a new understanding of the theoretical frameworks under study.

Although this paper attempts to draw attention to the benefits researchers can gain from using mixed models instead of GLM, mixed models are not without limitations. For instance, the mixed models analysis reported here used an unstructured covariance matrix. If the researcher opts to use a different covariance structure, the question naturally arises: Which covariance structure is most appropriate? While guidelines are available to help researchers choose the best-fitting structure (Dawson, Gennings, & Carter, 1997; Hamer & Simpson, 1999; SAS Institute, 1996; Wolfinger, 1993), the process can be challenging for the uninitiated.

Mixed models also have difficulty handling designs with multiple within-subjects factors. So-called doubly-multivariate repeated measures designs, in which multiple dependent variables are measured at multiple measurement occasions (Stevens, 1996) are not yet easily implemented in PROC MIXED. However, new software developments from SAS continue to bring mixed models procedures to the forefront, and future versions will likely include easily specified multiple within-subjects factor and doubly-multivariate designs in mixed models computing routines.

Despite several limitations, mixed models are still a better choice than the GLM for many research questions. These include the standard between by within-subjects ANOVA design described above, as well as growth models and repeated measures models with time-varying covariates. Based on the limitations of the GLM and the findings discussed above, the authors echo the recommendations of Wolfinger & Chang (1995) that researchers add the mixed models technique to their analytical toolbox in order to aid their search for answers to longitudinal research questions.

References

Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumackers (Eds.) Advanced structural equation modeling. Mahweh, NJ: Lawrence Erlbaum.

Campbell, A., Converse, P., & Rogers, W. (1976). The quality of American life: Receptions, evaluations, and satisfaction. New York: Russell Sage Foundation.

Dawson, K.S., Gennings, C., & Carter, W.H. (1997). Two graphical techniques useful in detecting correlation structure in repeated measures data. The American Statistician, 51(3), 275-283.

Duncan, T.E., Duncan, S.C., & Li, F. (1998). A comparison of model- and multiple imputation-based approaches to longitudinal analyses with partial missingness. Structural Equation Modeling, 5(1), 1-21.

Fairclough, D.L., Peterson, H.L., & Chang, V. (1998). Why are missing quality of life data a problem in clinical trials of cancer therapy? Statistics in Medicine, 17, 667-677.

Graham, J.W., Hofer, S.M., Donaldson, S.I., MacKinnon, D.P., & Schafer, J.L. (1997). Analysis with missing data in prevention research. In K.J. Bryant, M. Windle, & S.G. West (Eds.), The science of prevention, (pp. 325-366). Washington, DC: American Psychological Association Press.

Hamer, R. & Simpson, P.M. (1999). Mixed-up models: Things that look like they should work but don't, and things that look like they shouldn't work but do. Paper presented at the SAS User's Group International conference.

Harris, R.J. (1985). A primer of multivariate statistics. New York: Academic Press.

- Hays, W.L. (1994). Statistics. Fort Worth: Harcourt Brace College Publishers.
- Heitjan, D.F. (1997). Annotation: What can be done about missing data? Approaches to imputation. American Journal of Public Health, 87(4), 548-550.
- Hocking, R. R. (1996). Methods and applications of linear models. New York: John Wiley and Sons.
- Howell, D. C. (1987). Statistical methods for psychology. Boston, MA: PWS-Kent Publishing Company.
- Huston, T. L., Caughlin, J. P., Houts, R. M., Smith, S. E., & George, L. J. (in press). The connubial crucible: Newlywed years as predictors of marital delight, distress, and divorce. Journal of Personality and Social Psychology.
- Huston, T. L., & Houts, R. M. (1998). The psychological infrastructure of courtship and marriage: The role of personality and compatibility in romantic relationships. In T. Bradbury (Ed.), The developmental course of marital dysfunction. New York: Cambridge University Press.
- Littell, R.C., Milliken, G.A., Stroup, W.W., & Wolfinger, R.D. (1996). SAS system for mixed models. Cary, NC: SAS Institute, Inc.
- Little, J.W.A. & Rubin, D.A. (1987). Statistical analysis with missing data. New York: John Wiley and Sons.
- Roth, P. (1994). Missing data: A conceptual review for applied psychologists. Personnel Psychology, 47, 537-560.
- Searle, S. (1987). Linear models for unbalanced data. New York, NY: John Wiley and Sons.
- SAS Institute (1990). SAS/STAT user's guide, volume 1. Cary, NC: SAS Institute, Inc.

SAS Institute (1996). Advanced general linear models with an emphasis on mixed models. Cary, NC: SAS Institute, Inc.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics.

Smith, S., & Huston, T. (1998). Dimensions of early marital satisfaction: paths to distress and divorce. Presented at the 1998 International Conference on Personal Relationships, Saratoga Springs, New York.

Stevens, J. (1996). Applied multivariate statistics for the social sciences. Mahwah, N.J.: Lawrence Erlbaum Associates.

U.S. Bureau of the Census (1997). Statistical abstract of the United States: 1997 (117th edition). Washington, D. C.

Vangelisti, A. L., & Huston, T.L.(1994). Maintaining marital satisfaction and love. In L. S. Stafford & D.J. Canary (Eds.), Communication and Relational Maintenance (pp. 165-186). New York: Academic Press.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. Communications in Statistics—Simulations, 22(4), 1079-1106.

Wolfinger, R. & Chang, M. (1995). Comparing the SAS GLM and MIXED procedures for repeated measures. Orlando, FL: Proceedings of the North American SAS User's Group International Conference.

Author Notes

The authors would like to thank Ted L. Huston for providing data from the PAIR Project, which is supported by grants to Ted L. Huston from the National Science Foundation (SBR-9311846) and the National Institute of Mental Health (MH-33938). Correspondence concerning this article should be addressed to Dr. Tor Neilands, Academic Computing and Instructional Technology Services, COM 1/G2700, University of Texas at Austin, Austin, TX, 78712. E-mail correspondence may be directed to t.neilands@cc.utexas.edu.

Appendix A

The SAS PROC MIXED software syntax used to produce the results reported in the example appears below. Unlike GLM procedures that expect repeated measurements to reside in separate variables (one variable per measurement occasion), PROC MIXED expects a single dependent variable and a second variable indicating measurement occasion number. Thus, each case or observation has multiple rows or records of data for a PROC MIXED analysis with each row of data representing a single measurement occasion. Singer (1998) provides excellent examples of restructuring data from single-record to multiple-record format.

Beginning with SAS release 7.0, mixed models repeated measures analysis of variance is available via pull-down, point-and-click menus in the SAS Analyst application. To access the Analyst application, choose the Solutions menu, then Analysis, then Analyst. The Analyst application loads a spreadsheet into which the investigator can import data from SAS datasets or external files. The researcher can then select the Statistics menu, choose ANOVA, and then choose Repeated Measures.

The syntax shown below illustrates how to read an SPSS portable data file into SAS for Windows. The data are initially assumed to be in multivariate format where each measurement occasion is represented by a distinct dependent variable (e.g., cpsex1). A GLM analysis is illustrated. The program then converts the data structure from multivariate format to univariate format, in which each case has multiple records (one record per measurement occasion). Finally, the mixed models analysis is run, including the planned contrasts. See Littell et al. (1996) and SAS Institute (1996) for more details on specifying PROC MIXED syntax for repeated measures ANOVA models.

```

LIBNAME port SPSS 'f:\sexsat.por';

DATA one ;
    id+1 ;
    SET port._first_ ;
RUN ;

PROC GLM data = one ;
    TITLE 'GLM Analysis' ;
    CLASS mar4stt4 ;
    MODEL cpsex1 cpsex2 cpsex3 = mar4stt4 ;
    REPEATED time 3 / PRINTE ;
    LSMEANS mar4stt4 ;
RUN ;

DATA two ;
    SET one ;
    ARRAY score[3] cpsex1 cpsex2 cpsex3 ;
    DO phase = 1 to DIM(score) ;
        cpsex = score[phase] ;
        OUTPUT ;
    END ;
    KEEP id mar4stt4 phase cpsex ;
RUN ;

PROC MIXED DATA = two INFO IC UPDATE COVTEST ;
    TITLE 'Mixed Models Analysis' ;
    CLASS mar4stt4 phase id ;
    MODEL cpsex = mar4stt4 phase mar4stt4*phase ;
    REPEATED phase / SUBJECT=id TYPE=un GROUP=mar4stt4 ;
    LSMEANS mar4stt4*phase ;
    ESTIMATE 'grp Married Not Happy time 1 vs 3'
        phase 1 0 -1
        mar4stt4*phase 0 0 0 1 0 -1 0 0 0 0 0 0 / E;

    ESTIMATE 'grp Divorced Early time 1 vs 3'
        phase 1 0 -1
        mar4stt4*phase 0 0 0 0 0 0 1 0 -1 0 0 0 / E;
RUN ;

```