Behavior Genetics, Vol. 27, No. 2, 1997

The Use of Likelihood-Based Confidence Intervals in Genetic Models

Michael C. Neale[1–3] and Michael B. Miller[2]

Received 2 Mar. 1996—Final 10 Oct. 1996

This article describes the computation and relative merits of likelihood-based confidence intervals, compared to other measures of error in parameter estimates. Likelihood-based confidence intervals have the advantage of being asymmetric, which is often the case with structural equation models for genetically informative studies. We show how the package Mx provides confidence intervals for parameters and functions of parameters in the context of a simple additive genetic, common, and specific environment threshold model for binary data. Previously published contingency tables for major depression in adult female twins are used for illustration. The support for the model shows a marked skew as the additive genetic parameter is systematically varied from zero to one. The impact of allowing different prevalence rates in MZ vs. DZ twins is explored by fitting a model with separate threshold parameters and comparing the confidence intervals. Despite the improvement in fit of the different prevalences model, the confidence intervals on all parameters broaden, owing to their covariance.

KEY WORDS: Confidence intervals; structural equation modeling; Mx; matrix algebra; statistics; boundaries; nonlinear constraints; quantitative data; binary data; threshold models; covariance structure.

INTRODUCTION

Researchers in behavior genetics often use structural equation models (SEMs) in the analysis of twin or family data or both. Point estimates of parameters from these models, such as heritability estimates, may be reported along with information about their precision in the form of either confidence intervals (CIs) or standard errors (SEs), of which the latter can be used to produce CIs under the assumption of normally distributed estimators. The present paper discusses the merits of several CI approaches and focuses primarily on methods based on likelihood. We show that SEs often provide very poor approximations to CIs obtained by direct analysis of the log-likelihood. We describe new functionality that has been added to the SEM package Mx (Neale, 1995) for user-friendly, efficient computation of likelihood-based CIs for parameters and functions of parameters in a wide variety of models.

Maximum likelihood (ML) is the dominant form of estimation in SEM today. It is the default fit function in almost all the packages [Amos (Arbuckle, 1994), CALIS (SAS, 1988), EQS (Bentler, 1989), LISREL (Jöreskog and Sörbom, 1989), Mx (Neale, 1995)]. ML estimates have many desirable properties (see Silvey, 1975, for a general introduction to ML estimation); in particular, they are

(1) asymptotically unbiased,
(2) asymptotically efficient (uses all the data to estimate the parameters),
(3) sufficient (describes the likelihood up to an arbitrary constant), and

[1] Department of Psychiatry, Medical College of Virginia, Box 980126, Richmond, Virginia 23298. Telephone: (804) 828-3369. Fax: (804) 828-1471. E-mail: neale@psycho.psi.vcu.edu.
[2] Department of Psychiatry (Box 8134), Washington University School of Medicine, 4940 Children's Place, St. Louis, Missouri 63110.

0001-8244/97/0300-0113$12.50/0

(4) invariant (one-to-one transformations of parameters can be back transformed to recover the untransformed maximum),

and it can be argued (e.g., Edwards, 1972) that likelihood should be the basis for all parametric statistical inference. The general approach to interval estimation presented in this paper can be applied to any fit function, but it is most natural in the context of ML estimation because of its origin in likelihood theory.

Confidence Intervals

A confidence interval (CI) is determined by the position of its endpoints, called the upper and lower limits of the interval. A useful way to interpret CIs is through their relation to hypothesis tests; a $(1 - \alpha)$ CI consists of all those values of the parameter that cannot be rejected by an α-level hypothesis test. Suppose that the true population value of a parameter is θ and that we have collected appropriate data to obtain an estimate of it, $\hat{\theta}$. If we wish to test the hypothesis that $\theta = a$, where a is a possible value of θ, we can do so if there is an α-level test of this loss of fit available. We can use this test to construct a $(1 - \alpha)$ CI for θ by progressively moving the value of a away from $\hat{\theta}$ (and optimizing over the other parameters of the model) until we've determined the point where the hypothesis $\theta = a$ is rejected. In the case of MLEs this is straightforward, because the difference in fit is asymptotically distributed as χ^2 with one degree of freedom. When exact α-level statistical tests are not available, approximate tests often can be used to produce approximate $(1 - \alpha)$ CIs.

In statistical terms, we say that data are "random" when they cannot be predetermined because they are influenced by chance. Almost all data in behavioral genetic studies are random. Confidence intervals are determined by the data, so they too depend on random processes and therefore we refer to them as "random intervals." This leads to a definition:

> A $(1 - \alpha)$ CI for a parameter is a random interval that covers the *true* value of the parameter with probability greater than or equal to $(1 - \alpha)$.

In other words, we require a method for determining intervals from the data and the intervals so constructed must contain the true parameter value with probability $(1 - \alpha)$, no matter what the true value of the parameter. An ideal way to establish that a function is yielding a true $(1 - \alpha)$ CI would be from the joint distribution of the limits of the interval for every possible value of the parameter. Unfortunately, in most areas of statistical work, including structural equation models, it is not possible to determine the joint distribution of the limits of the CI, so approximations must be made.

Standard Errors Based on Derivatives

The conventional approach to estimating standard errors on parameters relies on the fact that the inverse of the information matrix (Fisher, 1922)—which reflects the curvature of the log-likelihood surface at its maximum—equals, in infinitely large samples, the variance–covariance matrix of parameter estimates. Thus, the square roots of the diagonal elements of the matrix inverse approximate the standard deviations of the parameter estimates. Conceptually, this is a very simple idea—to make use of a measure of the variance of the quantity in which we are interested. In practice, this matrix is readily available, as its inverse is required in many optimization methods (Gill *et al.*, 1991). Some optimizers use explicit derivatives of the likelihood function, which are usually accurate but may be expensive to compute. Other optimizers use numerical estimates of the derivatives, which are more convenient but less accurate (see Dolan and Molenaar (1991) for comparison of standard errors based on numerical vs. explicit derivatives). Neither method is ideal for the computation of confidence intervals or significance tests on parameters.

Despite its conceptual simplicity and speedy computation, the information matrix approach has three significant limitations which are described in the following paragraphs. All three of these limitations stem from the fact that the information-matrix method assumes implicitly that the log-likelihood is exactly quadratic in form. When the log-likelihood is quadratic, the information-matrix method and SEs will always give the same results as the method we advocate. Such perfect conformity of the log-likelihood to a quadratic form is rare in SEMs. For example, Fig. 2 shows the profile

log-likelihood for the additive-genetic parameter a in an SEM. The information matrix method would fit a parabola to the peak of this curve and would then treat the fitted parabola as if it were the profile log-likelihood itself. Obviously, the profile log-likelihood is not well fitted by a parabola and the results of the information-matrix method are extremely misleading in this example. Software programs that use the information matrix to compute SEs do not test the fit of the quadratic to the profile log-likelihood and so do not warn the user of even the most extreme problems.

First, the mean and variance of the statistic, even if estimated correctly, are not sufficient for producing confidence intervals. Interval estimation requires us to make a probability statement. Researchers typically apply an assumption of a normal distribution to form their interval and determine coverage probabilities. In large samples the distribution of an ML estimator tends to normality (Fisher, 1922) but "large" may be quite different from one model to another, as it depends on the statistical power of the design. When power is low, the distribution of a statistic may be skewed or kurtotic, so the mean and variance do not adequately describe the function. As we shall see, parameter estimates from genetic structural equation models can have highly asymmetric distributions, so conventional standard errors, and confidence intervals based upon them, are of limited use.

Second, the commonly used t-statistic to assess significance is not invariant to transformation. That is, a t-statistic based on parameter a is not the same as one based on the parameter a^2, even if the model is in all other respects equivalent (Neale *et al.*, 1989). This limitation is not restricted to significance testing; confidence intervals based on the standard error from the information matrix would not be invariant to transformation. Fortunately, the likelihood interval method described here does not suffer from these deficiencies. Venzon and Moolgavkar (1988) developed a method that is similar to ours but less flexible. They used the Newton–Raphson algorithm and analytical first and second derivatives of the log-likelihood. Their approach allows rapid convergence to the limits of the CI for any parameter in a model, but it does not allow for use of functions of parameters without rewriting the model in a different parameter space. Our method offers more flexibility because it computes the derivatives numerically and allows one to find CIs for functions of parameters without having to rewrite the model.

Third, parameters in genetic models often have bounds, either explicit or implicit. The assumed symmetry of the SE approach is inappropriate when parameter estimates are close to their bounds, which we can illustrate with two cases in the ACE model. Being a variance component, e^2 has a theoretical lower bound of zero. However, as e^2 approaches zero, the MZ twin covariance approaches the predicted phenotypic variance. A perfect covariance like this has likelihood of zero (the determinant of the predicted covariance matrix is zero) and log-likelihood of $-\infty$. In this case the confidence interval on e^2 does have an appropriate value near zero, which the likelihood method should obtain whereas the SE method typically provides a nonsensical negative lower limit for e^2.

COMPUTATION OF LIKELIHOOD-BASED INTERVALS

In an earlier paper, we noted that the likelihood-ratio (LR) statistic is invariant to the scaling of a parameter (Neale *et al.*, 1989). For example, consider the LR test of significance for heritability when we estimate a in a path model vs. when we fit a variance components model and estimate a^2 (see Fig. 1). The goodness-of-fit χ^2 will change by exactly the same amount if we fix a at zero in the path model or if we fix a^2 at zero in the variance components model. This may seem unsurprising, but Neale *et al.* (1989) showed that the information matrix t-statistic would give different answers for the two models.

When a parameter is moved away from its ML estimate, and fixed at another value, we can compare the likelihood under the two models. The model with the parameter fixed may be regarded as a submodel of the model in which it is free, so the usual likelihood ratio test (Bollen, 1989; Neale and Cardon, 1992) may be used to compare the models. Only one parameter is displaced—the remaining parameters are free to be reestimated—so the test has one degree of freedom. Twice the difference in the log-likelihood of the two models is distributed as χ^2, which provides a basis to assess the probability of finding results within a certain interval in successive experiments. Often reported are the 95%

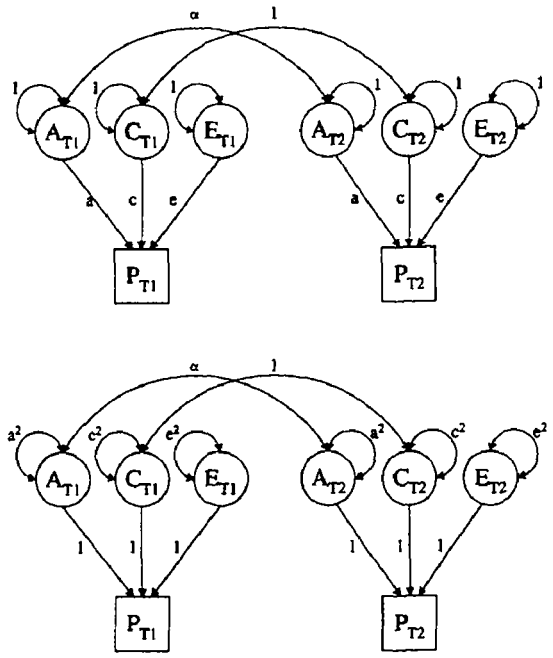

Fig. 1. Two parameterizations of a simple model for twin data. Additive genetic (A), common environment (C), and specific environment (E) components cause variation in the phenotypes of twins (*T*1 and *T*2). The top figure shows a standard path diagram; the bottom, a variance components model.

confidence intervals; these would be obtained by finding the point at which the χ^2_1 reaches the .05 level of significance ($\approx$3.84) in each direction. Although it might be thought that the two 5% tails (one in each direction) sum to give 10% of the distribution outside the interval and hence a 90% confidence interval, this is not the case. The χ^2 test in each direction is conditional on the value being on one side of the distribution, so it is still a 95% interval when both sides are considered jointly.

In practice, it is tedious to fix a parameter at a variety of locations to establish the confidence interval. To automate the procedure within Mx, we use the following procedure: let the ML solution be F_{ML}, then

(1) find the minimum of $F = -2F_{ML}$;
(2) Redefine the fit function to be $F_L = ((F + 3.84) - F_a)^2 + a$, where F_a is the fit under the revised model, and a is the parameter whose lower bound we seek;
(3) find the minimum of F_L to find the lower bound on a;
(4) reset the function to the solution of step 1;
(5) redefine the fit function to be $F_U = ((F + 3.84) - F_a)^2 - a$, where F_a is the fit under the revised model, and a is the parameter whose upper bound we seek; and
(6) find the minimum of F_U to obtain the upper bound on a.

At this point, the disadvantage of the interval approach is clear: it requires two further optimization steps for each set of confidence intervals on a parameter. However, computers are approximately doubling in speed each year and for many cases are fast enough to perform these steps in a reasonable amount of time. In addition, by starting at the solution of step 1, the minimizations in step 3 and 6 are relatively rapid. We agree with statisticians Meeker and Escobar (1995) in their assessment that "with improvements in computing technology, [computational cost] is less a problem today than in the past, and the direction for the future is clear; lack of appropriate easy-to-use software is the main problem" (p. 50). The new version of Mx includes an Interval option that makes computation of CIs very easy for the user.

Our algorithm is much faster than a systematic search. If such a search took only 10 steps to locate the boundary, it would be 10 optimizations over $p - 1$ parameters. Our method usually requires only one optimization over p parameters per limit. There is a problem because the solutions of the fit function in items 3 and 6 above are not exactly the boundary we seek but are biased by a constant that depends on the slope of the function at that point. At the minimum, the partial derivatives of the function with respect to the parameters should be zero, so in the case of parameter a, we have

$$\frac{\partial F_L}{\partial a} = -2\,[(F + 3.84) - F_a]\,\frac{\partial F_a}{\partial a} + 1 = 0$$

which can be rearranged to give

$$F_a = (F + 3.84) - \frac{1}{2\,\partial F_a / \partial a} \qquad (1)$$

For a steep slope the bias is small, but for shallow slopes there may be inaccurate estimation. To correct for this, Mx may reoptimize to improve the estimate. It is tempting to think that the error could be made arbitrarily small by multiplying the quadratic term $((F + 3.84) - Fa)^2$ by a constant, which

would have the effect of replacing the 2 by $2k$ in the denominator of equation 1. Unfortunately this is not practical when numerical estimates of the derivatives are being used. Too large a value of k will "freeze" the parameter at its starting point as the steepness of the quadratic function overpowers the optimization. Too small a value leads to bias in estimate of the bound. We can detect the latter condition by comparing the deviation of Fa from $F + 3.84$, and revising the value k appropriately. Although it is not a perfect algorithm, it is very practical and it works well in our test cases. Exact derivatives of the function would provide a better solution, but they would limit the general usefulness of the method.

The value of taking this approach in the Mx package is that confidence intervals may be computed on both parameters and (possibly nonlinear) functions of them. The fit function does not need to depend on the function of parameters in order to obtain confidence intervals on the functions of parameters. An example will make this clearer.

ILLUSTRATION: THE ACE MODEL

For illustration we turn to the simple additive genetic, common, and specific environment (ACE) model commonly used in the analysis of twin data. Rather than proceed with covariance matrices, we use the contingency tables published by Neale and Cardon (1992) for major depression assessed on a sample of adult female twins drawn from birth records in the state of Virginia, USA (Table I). The characteristics of the sample, the diagnostic measures used (DSM-IIIR) and zygosity determination have been described in detail elsewhere (Kendler *et al.*, 1992). A threshold model (Falconer, 1965) was fitted to the contingency table data via maximum likelihood, using the package Mx (Neale, 1995) and the script shown in the Appendix. The thresholds were constrained to be equal for twin 1 and twin 2 and across MZ and DZ twins. Table II shows parameter estimates, their 95% confidence intervals, and goodness of fit statistics. Table II also shows the estimates and confidence intervals for the parameters a and c for comparison with a^2 and c^2. As expected, the square of the upper limit on parameter a is equal (within rounding error) to the upper limit on parameter a^2, and the same relation holds for the other limits.

Table I. Contingency Tables of Twin Pair Diagnosis of Lifetime Major Depressive Illness

	MZ Twin 1		DZ Twin 1	
Twin 2	Normal	Depressed	Normal	Depressed
Normal	329	83	201	94
Depressed	95	83	82	63

Table II. Maximum-Likelihood Estimates (MLE), Confidence Intervals, and Goodness-of-Fit Statistics Obtained by Fitting Threshold Models 1 and 2 to the Data in Table I

	Model 1 $\chi^2_3 = 1.78$,			Model 2 $\chi^2_3 = 7.07$,		
		95% CI			95% CI	
Parameter	MLE	Lower	Upper	MLE	Lower	Upper
a^2	0.43	0.10	0.53	0.43	0.12	0.53
c^2	0.00	0.00	0.26	0.00	0.00	0.25
e^2	0.57	0.47	0.69	0.57	0.46	0.69
a	0.65	0.33	0.73	0.66	0.34	0.73
c	0.00	0.00	0.51	0.00	0.00	0.50
t_{MZ}	0.55	0.46	0.63	0.48	0.42	0.54
t_{DZ}	0.40	0.31	0.49	0.48	0.42	0.54

In Fig. 2, the support for the model is plotted as a function of a and a^2. We define support as the natural logarithm of the likelihood (Edwards, 1972) and we have rescaled it to have a maximum of zero in this figure. Immediately we see the marked asymmetry of the curves, which are much steeper on the right than the left-hand side of Fig. 2. The asymmetry reflects the fact that models with very high values of a are much more strongly rejected by the data than those with low values of a. High correlations have much smaller variances than low ones, so there is more power to reject the false model. We see the reflection of this asymmetry in the confidence intervals in Table II; the lower intervals are much farther from the estimate than the upper intervals.

The vertical lines in Fig. 2 illustrate the invariance to transformation property. A change in support of 2 units is shown by the horizontal line; the first place this line intersects the curve for the support for the model with a^2 parameterized is a lower limit on a, which is slightly less than .1 (first vertical line). The first intersection with the curve for a is the corresponding lower limit on a, which

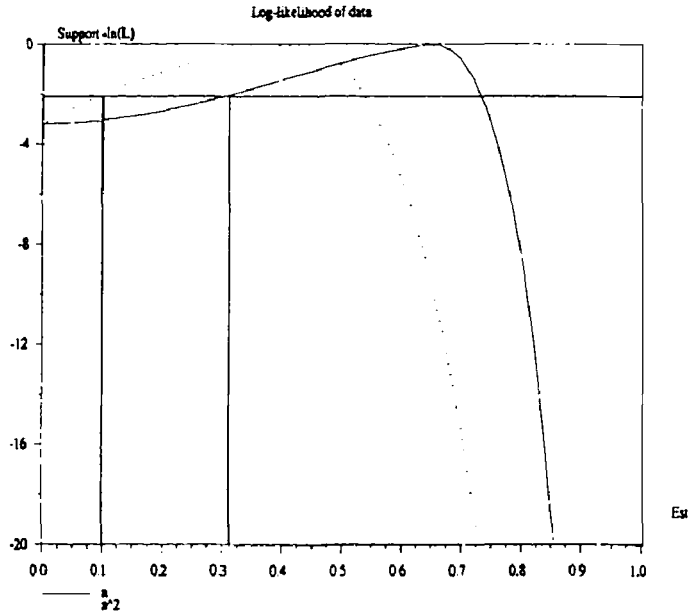

Fig. 2. Plot of support (ln(L)) against fixed values of parameter estimates a and a^2. The ln(L) is maximized with respect to all other parameters in the model for each value of a and a^2. The maximum of a^2 corresponds to the square of the maximum of a, illustrating the invariance property.

is slightly greater than .3, and is the square root of the limit on a^2.

Returning to Table II, the three right-hand columns show the estimates and confidence intervals when the thresholds are allowed to differ between MZ and DZ twins. As expected, the estimates of a, c and e are relatively close in the two models. However, there is a substantial improvement in fit when the thresholds are allowed to differ ($\chi^2_1 = 5.28$). The threshold for MZ twins is higher than that of DZ twins, indicating a lower rate of MDD in MZ twins in this sample. The confidence intervals are slightly broader under the model with different thresholds.

DISCUSSION

There is a general move away from significance tests in the social sciences (Savitz, 1993; Cohen, 1994). While their value is still a matter of debate, it is clear that decision making based on an arbitrary .05 significance level is not wise. Thus testing for significance of a^2 in a genetic model via a likelihood-ratio test gives only part of the information about the precision of this estimate. Confidence intervals give a better representation of what we have learned from our data. In this article, we showed an example of asymmetric confidence intervals on heritability, and illustrated their invariance to transformation. These factors favor the use of likelihood-based confidence intervals over those derived from standard errors extracted from the information matrix.

While likelihood-based confidence intervals are invariant to the equivalent reparameterization of a model (e.g., a vs. a^2), it is clear that they are not invariant to substantive changes in the model itself. If two parameters correlate, then fixing one of them will reduce the confidence interval on the other. The researcher must therefore avoid finding confidence intervals with methods that leave all other parameters fixed. While reoptimization to find the lower and upper bounds can be costly in terms of computer time, this seems a small price to pay for better information. The Interval feature in Mx 1.38 (http://griffin.vcu.edu/mx) is simple to use and minimizes this burden.

Caution is required as to the choice of model used to calculate confidence intervals. At this time, there is still debate over how to select a model. Consider, for example, eliminating the parameter c from the ACE model. Fixing c at zero reduces the

confidence intervals on correlated parameters such as a. In our example data, the estimate and confidence intervals of a^2 change from (.10,.53) to (.31,.53). If the "real world" or true model involved a small amount of c, we would have presented an unrealistically small confidence interval for a. Subsequent experimentation might deliver conflicting results, with a disproportionate number of parameter estimates outside say the 90% confidence interval. Superficially, it would seem that assuming that $c = 0$ is a foolish thing to do, because it increases the chances of being wrong. Yet scientific principles might make us take a different course of action. Falsifiability is a valuable quality in scientific theories (Popper, 1961). The fact that our model with no c is more readily falsified should be viewed as a positive attribute.

Although the preceding argument was posed in terms of c being zero, it might just have easily been expressed in terms of a being zero. Two factors influenced the choice of $c = 0$: (i) empirically c often is small (Rowe, 1994), and (ii) we have reason to expect that c may diminish over time in adult populations. If effects of the environment decay over time (many plausible differential equation models would predict this), then adult twins who no longer share significant shared environmental effects with their cotwin are likely to grow less similar (though see Hopper and Culross, 1983). In the limit, the effect of their shared rearing experience could be zero. In contrast, genetic effects have a continuous opportunity to make the twins similar, regardless of their geographical circumstances ($G \times E$ notwithstanding). Obviously, we would need to control for adult cohabitation effects (Kendler *et al.*, 1993) to partition shared rearing experiences from shared adult environment.

Finally, there are some limitations to consider. First, minus twice the log-likelihood only approaches χ^2 asymptotically under certain conditions (Bollen, 1989). For small sample sizes it may be appreciably different from the χ^2 distribution, so confidence intervals based upon it may be inaccurate. In this case we might resort to an alternative fit function, such as minimum χ^2 (Agresti, 1990). Second, bounds can present interpretational problems for the computation of confidence intervals, when the lower theoretical bound on a parameter does not yield the given decrease in fit required for a certain level of confidence (e.g., 3.84 for 95%). Reporting the lower bound seems appropriate here, but it should be noted that this is a boundary value. Third, it is possible that the log-likelihood is nonmonotonic such that more than one point on either side of the optimum will appear to meet criteria for an upper or lower limit of the CI. For example, minus the upper confidence interval on a also produces a lower CI, although reasonable models would constrain a to be nonnegative. Ideally an algorithm should be always give the limits that are within the parameter space but farthest on either side of the global optimum. The algorithm we recommend is not able to solve this problem, but researchers who are concerned about this possibility can plot the profile log-likelihood function to determine if the problem of local optima exists in their data. This can be accomplished in Mx by producing intervals of varying sizes and plotting the values of the chi-square statistic on the corresponding values of the upper or lower limit.

APPENDIX

Mx script to fit threshold model to MZ and DZ twin data and to obtain confidence intervals on a^2, c^2, e^2, a, c, and e.

```
!
! Script for threshold model
! MDD data, Kendler et al, 1992
!
G1: Model parameters
Calc NGroups=3
Begin Matrices;
I Ident 1 1
X Lower 1 1 Free
Y Lower 1 1 Free
W Lower 1 1
T Full 2 1
End Matrices;
! parameters are fixed by default, unless declared
! free
Begin Algebra;
A= X*X';
C= Y*Y';
D= W*W';
E= I-A-C-D;
End Algebra;
Specify T 3 3 ! put two new, equated parameters
! in T

End Group;
```

```
G2: Female MZ twin pairs
Data NInputvars=2
Labels dep  t1 dep-t2
CTable 2 2
329 83
95 83
Matrices= Group 1
Thresholds T /
Covariances A+C+D+E | A+C+D
            A+C+D | A+C+D+E
Options RSidual
End Group;

G3: Female DZ twin pairs
Data NInput  vars=2
Labels dep  t1 dep-t2
CTable 2 2
201 94
82 63
Matrices = Group 1
H Full 1 1
Q Full 1 1
Thresholds T /
Covariances A+C+D+E | H@A+C+Q@D
            H@A+C+Q@D | A+C+D+E /

Matrix H .5
Matrix Q .25
Start .6 All
Bound .001 1 Y 1 1 X 1 1
Intervals A 1 1 1 C 1 1 1 E 1 1 1 X 1 1 1 Y 1 1 1
End Group;
```

ACKNOWLEDGMENTS

This research was supported by NIH Grants RR-08123, MH/AA-49492, MH17104, and AA-09095.

REFERENCES

Arbuckle, J. (1994). *AMOS 3.51 Documentation Package*. Smallwaters Inc., Chicago.

Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.

Bentler, P. M. (1989). *EQS: Structural Equations Program Manual*, BMDP Statistical Software, Los Angeles.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*, John Wiley, New York.

Cohen, J. (1994). The earth is round. ($P < .05$) *Am. Psychol.* **49**:997–1003.

Dolan, C. V., and Molenaar, P. C. M. (1991). A comparison of four methods of calculating standard errors of maximum likelihood estimates in the analysis of covariance structure. *Br. J. Math. Stat. Psychol.* **44**:359–368.

Edwards, A. W. F. (1972). *Likelihood*, Cambridge, London.

Falconer, D. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**:51–76.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. London Ser. A* **222**:309–368.

Gill, P. E., Murray, W., and Wright, M. H. (1991). *Numerical Linear Algebra and Optimization, Vol. 1*, Addison-Wesley, New York.

Hopper, J. L., and Culross, P. R. (1983). Covariation between family members as a function of cohabitation history. *Behav. Genet.* **13**:459–471.

Jöreskog, K. G., and Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*, 2nd ed., SPSS, Inc., Chicago.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992). A population based twin study of major depression in women: Varying definitions of illness. *Arch. Gen. Psychiatry* **49**:257–266.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behav. Genet.* **23**:21–27.

Meeker, W. Q., and Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.* **49**(1):48–53.

Neale, M. C. (1995). *Mx: Statistical Modeling*, 3rd ed., Box 980126 MCV, Richmond VA 23298.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*, Kluwer Academic, Dordrecht.

Neale, M. C., Heath, A. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989). Fitting genetic models with LISREL: Hypothesis testing. *Behav. Genet.* **19**:37–69.

Popper, K. R. (1961). *The Logic of Scientific Discovery*, Science Editions (translation; original work published 1934).

Rowe, D. C. (1994). *The Limits of Family Influence: Genes, Experience, and Behavior*, Guilford Press, New York.

SAS (1988). *SAS/STAT User's Guide: Release 6.03*, SAS Institute, Inc., Cary, NC.

Savitz, D. A. (1993). Is statistical significance testing useful in interpreting data? *Reprod Toxicol.* **7**:95–100.

Silvey, S. D. (1975). *Statistical Inference*, Chapman and Hall, London.

Venzon, D. J., and Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *ApplStat, 3.*

Edited by Kay Phillips