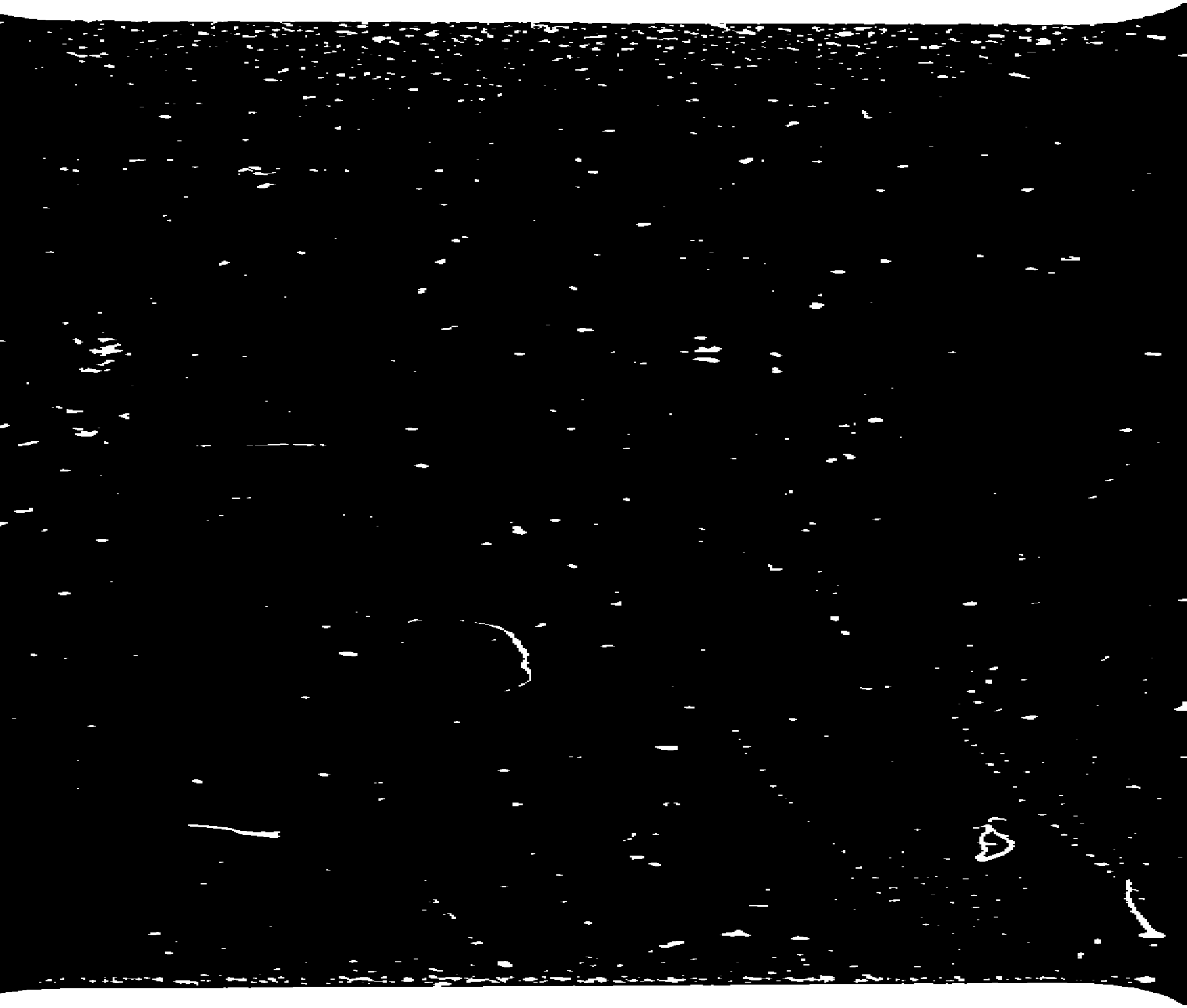
DOCUMENT RESUME

ED 199 282		TM 810 191

AUTHOR	McKinley, Robert L.; Reckase, Mark D.
TITLE	A Comparison of the ANCILLES and LOGIST Parameter Estimation Procedures for the Three-Parameter Logistic Model Using Goodness of Fit as a Criterion.
INSTITUTION	Missouri Univ., Columbia.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO	RR-80-2
PUB DATE	Dec 80
CONTRACT	N00014-77-C-0097
NOTE	88p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Comparative Analysis: *Computer Programs: Difficulty Level: *Goodness of Fit: Latent Trait Theory: *Mathematical Models: Maximum Likelihood Statistics: *Statistical Significance: Test Items
IDENTIFIERS	ANCILLES Estimation Procedures: Chi Square Test: Iowa Tests of Educational Development: Item Discrimination (Tests): LOGIST Estimation Procedures: Mean Square Fit: Meta Analysis: Three Parameter Model

ABSTRACT

A study was conducted to compare the quality of the item parameter estimates obtained from the ANCILLES and LOGIST estimation procedures using goodness of fit as a criterion. Statistics used to compare the fit included a chi-square statistic and a mean square deviation statistic. Other analyses performed included comparisons of the distributions of the parameter estimates obtained from the procedures, and a set of meta-analyses performed on the chi-square statistics obtained for the items. The data for the study were composed of 50 items and 2,000 cases obtained using a stratified random sample of 357 items and 4,000 cases of the Iowa Tests of Educational Development. Results indicated that there are qualitative differences in the estimates obtained from these two procedures. While the parameter estimate distributions obtained from these two procedures were similar, lack of fit occurred for significantly more items for ANCILLES than for LOGIST. Lack of fit for ANCILLES appeared to be stongly related to item difficulty, while for LOGIST it was related to item discrimination. Although LOGIST is more expensive to use than ANCILLES, ANCILLES yielded lack of fit significantly more often than LOGIST, and did not yield item parameter estimates for two items. (Author/RL)



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 80-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Comparison of the ANCILLES and LOGIST Parameter Estimation Procedures for the Three-Parameter Logistic Model Using Goodness of Fit as a Criterion		5. TYPE OF REPORT & PERIOD COVERED Technical report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Robert L. McKinley and Mark D. Reckase		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0097
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Educational Psychology University of Missouri Columbia, Missouri 65201		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N Proj.: RR042-04 T.A.:042-04-01 W.V.:NR150-395
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE December, 1980 13. NUMBER OF PAGES 76
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approval for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Latent Trait Models
ANCILLES
LOGIST
Goodness of fit
Parameter Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A study was conducted to compare the quality of the item parameter estimates obtained from the ANCILLES and LOGIST estimation procedures using goodness of fit as a criterion. Statistics used to compare the fit yielded by the procedures included a chi-square statistic and a mean square deviation statistic. Other analyses performed included comparisons of the distributions

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

of the parameter estimates obtained from the procedures, and a set of meta-analyses performed on the chi-square statistics obtained for the items. The data for the study was composed of 50 items and 2,000 cases obtained using a stratified random sample of 357 items and 4,000 cases of the Iowa Tests of Educational Development. Results indicated that differences did exist in the quality of parameter estimates obtained from the ANCILLES and LOGIST procedures. More items for ANCILLES showed significant lack of fit than for LOGIST, based on the chi-square tests. However, a test to determine whether the ANCILLES chi-squares were larger than the LOGIST chi-squares significantly more than half the time was not significant. Also, the dependent t-computed for the MSD statistics was not significant. Further analyses indicated that the parameter estimates obtained from the two procedures were highly correlated. Comparisons of the item parameter estimates with the chi-square values indicated that ANCILLES yielded poor fit for items with extreme difficulty values. Because of this, and the fact that ANCILLES did not provide estimates for all the items, it was concluded that there were real differences in the quality of the parameter estimates obtained for the two procedures.

CONTENTS

Introduction.. 1

The Model and Procedures.. 2

Goodness of Fit Statistics.. 3

Method.. 6
Test Data... 6

Analyses.. 6

Results... 7
Chi-Square Analyses... 7

MSD Statistics..58

Parameter Estimate Distribution Analyses..............................62

Item Parameter Estimates..62

Ability Estimates...65

Discussion..70
Chi-Square Analyses...71

MSD Statistics..72

Item Parameter Estimates..72

Ability Estimates...74

Summary and Conclusions...74

References..75

A Comparison of the ANCILLES and LOGIST Parameter Estimation Procedures for the Three-Parameter Logistic Model Using Goodness of Fit as a Criterion

Due to the growing use of latent trait models and the wide range of applications of these models (see the Journal of Educational Measurement, Summer, 1977), it has become important to investigate the properties of the numerous procedures that are available for estimating the parameters of the models. There are a number of different models in current use (e.g., one-, two-, and three-parameter logistic; graded response; nominal response), and for many of these models item parameters can be estimated in several ways. While there has been some research done to investigate the differences between the models (Reckase, 1977; Yen, in press; Divgi, 1980; Urry, 1970, 1977a), little has been done to compare estimation procedures for given model.

One commonly used latent trait model is the three-parameter logistic (3PL) model. There are at least three estimation procedures available for the 3PL model, each based on a different computer program. For example, the ANCILLES (Urry, 1978), OGIVIA (Urry, 1977b), and LOGIST (Wood, Wingersky, and Lord, 1976) programs are all designed to estimate parameters for the 3PL model. Very little has been done to study the differences in these three procedures. Although they are based on the same model, the methods that these programs employ to estimate the parameters for the model are quite different (the differences between ANCILLES and OGIVIA are not as great as the differences between LOGIST and the others). The few studies that have dealt with the differences in these procedures have primarily been concerned with the ability of the procedures to faithfully reproduce true item and ability parameters. For instance, in a simulation study conducted by Ree (1979), three groups of 2,000 subjects were simulated, and the simulated responses were calibrated using the ANCILLES, OGIVIA, and LOGIST procedures. The estimated parameters were compared to the true parameters, the estimated true scores, and an information comparison was made. It was concluded that the selection of an item calibration program should be dependent on the distribution of ability in the calibration sample, the intended use of the parameter estimates, and computer resources available. Specifically, the differences that were found included the finding that LOGIST performed best for rectangular ability distributions and OGIVIA performed best for normally distributed ability groups. Also, LOGIST was more expensive to run, but the OGIVIA and ANCILLES programs did not always give estimates for every item.

The Ree study indicated that there were differences in the quality of the parameter estimates given by the procedures considered, and the conclusions provided guidelines for selecting procedures for the model. This type of study is useful, and should be extended to include other models, but there are other comparisons that should be made. One important comparison that was not made was a comparison of the procedures using the fit of the model to the data as a criterion, an important factor when considering the quality of parameter estimation using the procedures. The purpose of this study, then, is to extend the comparison of the 3PL parameter estimation procedures

to include a comparison of the fit of the 3PL model to real data when using the different procedures. Before reporting the present study, however, a discussion of the model and procedures, as well as the fit statistics used, will be given.

The Model and Procedures

The model that was employed in this study was the three-parameter logistic model presented by Birnbaum (1968). The model requires three parameters for each item and one ability parameter for each examinee. The model is given by

$$P_i(\theta_j) = c_i + (1 - c_i) \frac{\exp(Da_i(\theta_j - b_i))}{1 + \exp(Da_i(\theta_j - b_i))} \tag{1}$$

where θ_j is the ability parameter for Examinee j, a_i is the item discrimination parameter, b_i is the item difficulty parameter, c_i is the item guessing parameter, $P_i(\theta_j)$ is the probability of a correct response to Item i, and D is a scaling constant equal to 1.7.

There are three commonly used programs for the estimation of the parameters of the 3PL model, ---ANCILLES, LOGIST, and OGIVIA---but because ANCILLES is a newer version of OGIVIA, OGIVIA was not included in this study. The ANCILLES estimation procedure is a two-staged procedure. In the first stage raw scores, corrected to exclude scores on the item being calibrated, are used as a measure of manifest ability. Using the correct raw scores the program computes item characteristic curves (ICC's) for various sets of guessing, discrimination, and difficulty values. The proportions of examinees falling within set intervals of the manifest ability who passed the item are computed, and those values are compared to the generated ICC's. Chi-square fit statistics are computed for each ICC, and the set of values with the minimum chi-square is selected. This procedure is repeated for all the items to be calibrated. Then a second stage is begun, in which unregressed Bayesian modal estimates (UBME's) are used as manifest ability in place of raw scores. This substitution is made because the UBME's more closely approximate the latent ability distribution. Using the UBME's ancillary estimates of the item parameters are made.

The LOGIST procedure, on the other hand, uses neither Bayesian modal ability estimation nor minimum chi-square item parameter estimation procedures. Rather, LOGIST uses maximum likelihood estimation for estimating both ability and item parameters. Initial values for the item parameter estimates are set, and ability estimates are computed for all the examinees using maximum likelihood estimation. Then the ability estimates are held fixed, and new estimates are made for the *a*- and *b*-values, again using maximum likelihood estimation. These two steps, called a stage, are repeated a number of times with the *c*-values held fixed.

After the first few stages the c-values are allowed to vary, but change in the c-values is still restricted. The procedure cycles through as many stages as is necessary to converge. Convergence is reached when the difference between the estimates for successive stages is less than errors of calculation.

While this is not a complete discussion of how these two procedures operate, it is clear from this treatment of the ANCILLES and LOGIST procedures that they do differ in the way in which parameters are estimated. For a more detailed discussion of these procedures see: Wood, Wingersky, and Lord, 1976; and Urry, 1978.

Goodness of Fit Statistics

Whenever a model is used to approximate real data it is important to determine the accuracy of the approximation. The failure of a model to accurately represent the data may result in inaccuracies in measurements based on that model. Goodness of fit of the model to empirical data, then, is clearly an important property to consider when selecting a model. It is just as important when considering which procedure to use for estimating the parameters of a model. Item parameter estimates for a model are not unique. Clearly, different procedures may result in different estimates for the same data. If different sets of estimates fit the data equally well, then either procedure may be appropriate. However, if the two sets of estimates do not fit the data equally well, the procedure yielding the best fit is the more desirable procedure.

In the past a number of statistical goodness of fit tests for gauging the fit of a model to data have been proposed. Generally, most of these tests involve computing statistics that fall in a chi-square or an approximate chi-square distribution. For instance, a fit statistic for the 1PL model proposed by Wright and Panchapakesan (1969) involves dividing examinees into groups according to number-right scores, and for each score group computing the observed and expected proportions of examinees passing the item, with the expected proportion being computed from the model. From these proportions a fit statistic is computed with the following formula:

$$\chi^2 = \sum_{\substack{j=1 \\ N_j \neq 0}}^{J} \frac{N_j(O_{ij} - E_{ij})^2}{E_{ij}(1 - E_{ij})} \qquad (2)$$

where O_{ij} is the observed proportion passing Item i in Score Group j, E_{ij} is the proportion predicted by the model, and the summation is over all score groups for which the number of examinees in the group is not zero. The summation over number-right score groups can be used since the number-right score is a sufficient statistic for estimating θ for the 1PL model. That is, each score group contains examinees with the same θ. This statistic is essentially the summation of squared z-scores. Wright and

Panchapakesan (1969) state that this statistic has J-1 degrees of freedom, where $N_j \neq 0$. A variation on this statistic used by Rentz and Bashaw (1975), involves computing the χ^2 above and then dividing it by the number of score groups for which the number of examinees in the group is not zero, obtaining as a result a 'mean square' fit statistic.

A procedure not limited to the 1PL model was proposed by Yen (in press). This statistic differs from the Wright and Panchapakesan statistic in that examinees are not grouped by number-right scores. Rather, examinees are ordered according to their ability estimates. The range of ability estimates is then divided into categories (Yen suggests 10), and the observed and expected proportions are computed for those categories. This fit statistic is given by

$$\chi^2 = \sum_{j=1}^{10} \frac{N_j(O_{ij} - E_{ij})^2}{E_{ij}(1 - E_{ij})} \tag{3}$$

where O_{ij} and E_{ij} are as defined previously. Since the categories for this statistic are not based on number-right scores this statistic is not limited to the 1PL model. Yen (in press) suggest that this statistic has 10-m degrees of freedom, where m is the number of item parameters estimated.

A similar statistic, s, was suggested by Wright and Mead (1977). This statistic is given by

$$s = \frac{1}{J} \sum_{j=1}^{J} \frac{N_j(O_{ij} - E_{ij})^2}{E_{ij}(1 - E_{ij}) - \sigma^2 p_j} \tag{4}$$

where O_{ij} and E_{ij} are as defined above and $\sigma^2 p_j$ is the variance within category j of the predicted proportions passing the item (Yen, in press). Wright and Mead suggest the addition of the $\sigma^2 p_j$ term because examinees within a category do not have the same θ, and the addition of the term provides a more accurate estimate of the variance of O_{ij} than does the denominator in Equation 3. For this statistic examinees are grouped in the same way as for the Yen statistic. However, Wright and Mead suggest six or fewer categories, rather than the 10 suggested by Yen. The constant 1/J provides a mean fit statistic for the J cateogires.

One statistic for measuring goodness of fit of a model to data that is not based on the chi-square distribution is the mean square deviation (MSD) statistic proposed by Reckase (1977). The MSD statistic is given by

$$MSD_i = \frac{\sum_{j=1}^{N} (u_{ij} - P_{ij})^2}{N}$$

where u_{ij} is the response to Item i by Examinee j, P_{ij} is the probability of a correct response as given by the model, and N is the number of examinees. The purpose of this statistic is to avoid the differences caused by different interval sizes encountered with the χ^2 statistics described above. Reckase suggests that, even though the sampling distribution of the statistic in unknown, hypotheses can still be tested because only comparative information is of interest. Thus, differences in MSD statistics obtained for different procedures for a single set of items can be tested using analysis of variance procedures (or, in the case of two procedures, a simple dependent t-test). Because the statistic does not group examinees, its use is not limited to a single model.

For the present study the MSD statistic and the chi-square statistic suggested by Yen were selected. Since the present study is concerned with procedures for estimating the parameters of the 3PL model, those fit statistics based on the number-right score groups are clearly inappropriate. In a comparison of the Yen statistic and the statistic proposed by Wright and Mead, Yen (in press) found virtually no difference in the two statistics. Yen concluded that using 10 categories was sufficient to produce small enough values of $\sigma^2 P_j$ would be sufficiently small so as to make it unnecessary to adjust the denominator in the chi-square statistic. Because of the concern over the differences the category sizes make in the chi-square statistic, the MSD statistic was included in the analyses.

Analyses for the current study, then, include the comparison of the chi-squares obtained for the two procedures using the statistic proposed by Yen, and a comparison of the MSD statistics obtained for the two procedures. In addition, direct comparisons of the obtained parameter estimates will be made. These comparisons will include descriptive statistics and correlations of the distributions of ability and item parameter estimates obtained from the ANCILLES and LOGIST programs, as well as plots of the observed proportions of examinees passing an item with the proportions predicted by the model using the estimates from the procedures.

Method

Test Data

The data-set used for this study was constructed from a 4,000 case sample of the Iowa Test of Educational Development (ITED). The test items were a stratified random sample of 50 items from the various subtests of the ITED. Response data for 1,999 examinees were sampled from among four grade levels.

Analyses

To begin the analyses, ability and item parameter estimates for the 3PL model were obtained by running both the ANCILLES and LOGIST calibration programs on the response data. For each set of estimates obtained chi-squares were computed for each of the items using the following procedure. First the range of ability estimates was divided into 49 categories of .1 width (the end categories were larger so as to keep all cell frequencies ≥ 5). Examinees were grouped, then, according to which category their ability estimates were in. For each category both the proportion of examinees in that category passing the item and the proportion failing the item were obtained. Also, for each category the expected proportion passing and the expected proportion failing the item were computed. The expected proportion passing an item, as predicted by the 3PL model, is

$$E_{ij} = c_i + (1 - c_i) \frac{\exp(1.7a_i(\theta_j - b_i))}{1 + \exp(1.7a_i(\theta_j - b_i))} \quad (6)$$

where E_{ij} is the proportion of examinees in Category j expected to pass Item i, θ_j is the midpoint of Category j, and the other parameters are as defined for Equation 1. It should be noted at this point that, due to the small category size, the variance of the expected proportions was quite small. For the purposes of this study, then, the expected proportions were assumed to be constant within a category. That is, the variance of the expected proportions is equal to zero.

Once the observed and expected proportions were obtained for both sets of parameter estimates, then chi-square statistics for each item, using both sets of estimates, were computed using Equation 3 (with the modification that 48 categories were used instead of 10). Using these chi-squares a number of analyses were performed. First, the chi-square values were compared to the critical value to determine whether they were significant. Then a comparison was made to determine which procedure resulted in lack of fit for more items. Then the chi-squares for each procedure were summed and the resulting chi-squares were tested for significant lack of fit for the test as a whole. Further analysis included performing a binomial test to determine whether the chi-squares obtained for one procedure were larger than the chi-squares obtained for the other procedure more times than would be expected by chance. Two final analyses using the chi-squares involved the graphic presentation of the obtained

values. One analysis involved plotting, for each category, the observed and expected proportions passing the item. This was essentially a visual comparison of empirical and theoretical ICC's for each item. Plots were made for both procedures. The last analysis performed with the chi-squares for each procedure was the plotting of the obtained distribution of chi-squares with the actual distribution of chi-squares computed from the chi-square probability density function.

A set of analyses was also performed using the MSD statistic set out in Equation 5. For each set of estimates MSD statistics were computed for each item. The resulting statistics were tested for significant differences using a dependent t-test.

A final set of analyses involved the direct comparison of the parameter estimates obtained from the ANCILLES and LOGIST procedures. The analysis included a comparison of the shape of the distributions of the ability and item parameter estimates, as well as correlations of the two sets of estimates.

Results

Chi-Square Analyses

The item chi-square statistics obtained for the ANCILLES and LOGIST procedures are presented in Table 1. Item 1 and Item 9 were deleted by ANCILLES during calibration. Comparison of these values to the critical value required for significance at $\alpha = .05$ revealed that significant lack of fit occurred for fifteen items for the ANCILLES procedure, and for six items for the LOGIST procedure. Although it is true that such a multiple comparison increases the probability of finding significant results, the intent is to compare the two procedures rather than to make an evaluation of the procedures across items. Therefore the alpha level was not adjusted to accommodate the multiple comparison. A test for the significance of the difference between two correlated proportions (Ferguson, 1976) yielded a $z = 2.68$, indicating that a significantly higher proportion of items showed lack of fit for the ANCILLES procedure than for the LOGIST procedure ($p < .05$).

Considering the results reported above it is somewhat surprising that the ANCILLES chi-square values are not larger than the LOGIST chi-square values for significantly more than half the items. The ANCILLES chi-square value is larger than the LOGIST chi-square value for only 25 items, and the ANCILLES mean chi-square was not significantly larger than the mean chi-square value for LOGIST (58.12 for ANCILLES and 52.44 for LOGIST). It would appear, then, that the ANCILLES chi-square values were not larger than the LOGIST chi-square values more often than would be expected by chance, but when they were larger than the LOGIST values, they tended to be significant.

Table 1

ANCILLES vs. LOGIST Goodness of Fit Comparison
Using Yen's Chi-Square Statistic

Item	ANCILLES	LOGIST
1		46.65
2	70.36*	50.88
3	58.95	48.49
4	39.95	43.73
5	116.45*	46.57
6	133.14*	50.72
7	34.00	41.97
8	41.83	61.64
9		46.04
10	56.56	32.90
11	51.13	48.37
12	46.97	38.58
13	81.97*	53.90
14	35.38	59.62
15	51.01	60.64
16	61.68*	52.02
17	75.22*	62.06*
18	62.22*	44.90
19	50.15	35.13
20	36.33	53.92
21	50.91	58.81
22	57.51	69.78*
23	104.66*	80.91*
24	45.96	46.93
25	48.84	48.26
26	52.44	55.91
27	93.87*	93.96*
28	57.10	56.14
29	76.61*	51.09
30	43.76	52.43
31	50.85	49.82
32	33.92	45.20
33	65.78*	58.18
34	52.86	70.37*
35	58.27	60.90
36	41.66	44.55
37	55.20	47.50
38	50.97	51.54
39	34.49	44.98
40	50.54	57.05
41	46.54	47.28
42	72.42*	71.08*
43	46.98	43.29
44	70.49*	39.81
45	62.56*	49.21
46	67.57*	50.89
47	28.85	38.22
48	45.55	46.20
49	55.85	56.35
50	53.39	44.24

Note. The critical value for rejection of adequate fit is $\chi^2(45) > 61.66$ at $\alpha = .05$.

* significant at .05 level.

Another analysis that was performed on the chi-square values obtained for the ANCILLES and LOGIST procedures was the summation of the chi-squares over items to test whether there was significant lack of fit for the test as a whole. Using the normal approximation to the chi-square distribution yields a standard deviation of 66. The ANCILLES chi-squares summed to 2789, which yielded a $\underline{z} = 9.54$. The LOGIST chi-squares summed to 2517, which resulted in $\underline{z} = 5.41$. Comparing these $\underline{z}$-score values to the standard normal distribution, clearly both summed chi-squares were significant, indicating that there was significant lack of fit for the test as a whole for both procedures.

The final analyses performed on the obtained chi-square values involved comparing the chi-square values to a graphic display of the empirical and theoretical plots of the item characteristic curves. Figure 1 through Figure 48 show the obtained and predicted proportions correct for each item plotted against the ability estimates. Plots were made for both the ANCILLES and LOGIST parameter estimates. Examining these figures closely does reveal one consistent pattern across items. The poorest fit for both procedures occurs at the lower end of the ability scale. This is not surprising since it was already known that the lower asymptote of the ICC is difficult to estimate. It should be noted, however, that the values at the lower end of the ability scale are somewhat distorted due to the collapsing of categories that was required for the chi-square procedure. In order to keep category frequencies above five, the collapsing of end categories was necessary, which resulted in some category frequencies that were relatively large due to the width of the category.

Using a visual comparison of the plots for the two procedures, it is difficult to determine whether the fit of one procedure was any better than the fit for the other procedure. It is also difficult to predict from the plots for which items lack of fit was significant. For example, the ANCILLES chi-square value for Item 6 was 133.14, while the LOGIST chi-square value for Item 6 was 50.72. The plots for Item 6, shown in Figure 5, do not at first indicate the large difference in fit. However, closer investigation does yield some insight as to cause of the difference in fit for that item. The intervals for the ANCILLES procedures showing the largest discrepancy between the observed proportion correct and the expected proportion correct are those intervals containing the greatest number of examinees. For instance, the intervals between $\theta = 1.0$ and $\theta = 2.0$ show a fair amount of discrepancy between the observed and expected proportions correct. In those intervals frequencies vary from 60 to 90 examinees. (see Figure 51). For the LOGIST procedure the poorest fit appears to occur near $\theta = 2.0$ and $\theta = -2.0$. Frequencies in those intervals range from 10 to 20 examinees, which is far lower than the frequencies in the intervals where the ANCILLES procedure showed poor fit. This was not a consistent pattern across items, however.

Figure 15 shows the plots for Item 17. Both procedures showed lack of fit for Item 17, and it appears from the plots that the poorest fit was in the same ability ranges for both procedures. For Item 23, shown in Figure 21, the ANCILLES procedure shows lack of fit in approximately the same ability ranges as in other items discussed, but the LOGIST procedure appears to fit poorly across the entire ability range. The plots, then, do not appear to indicate any other consistent pattern for the procedures.

FIGURE 1

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 2

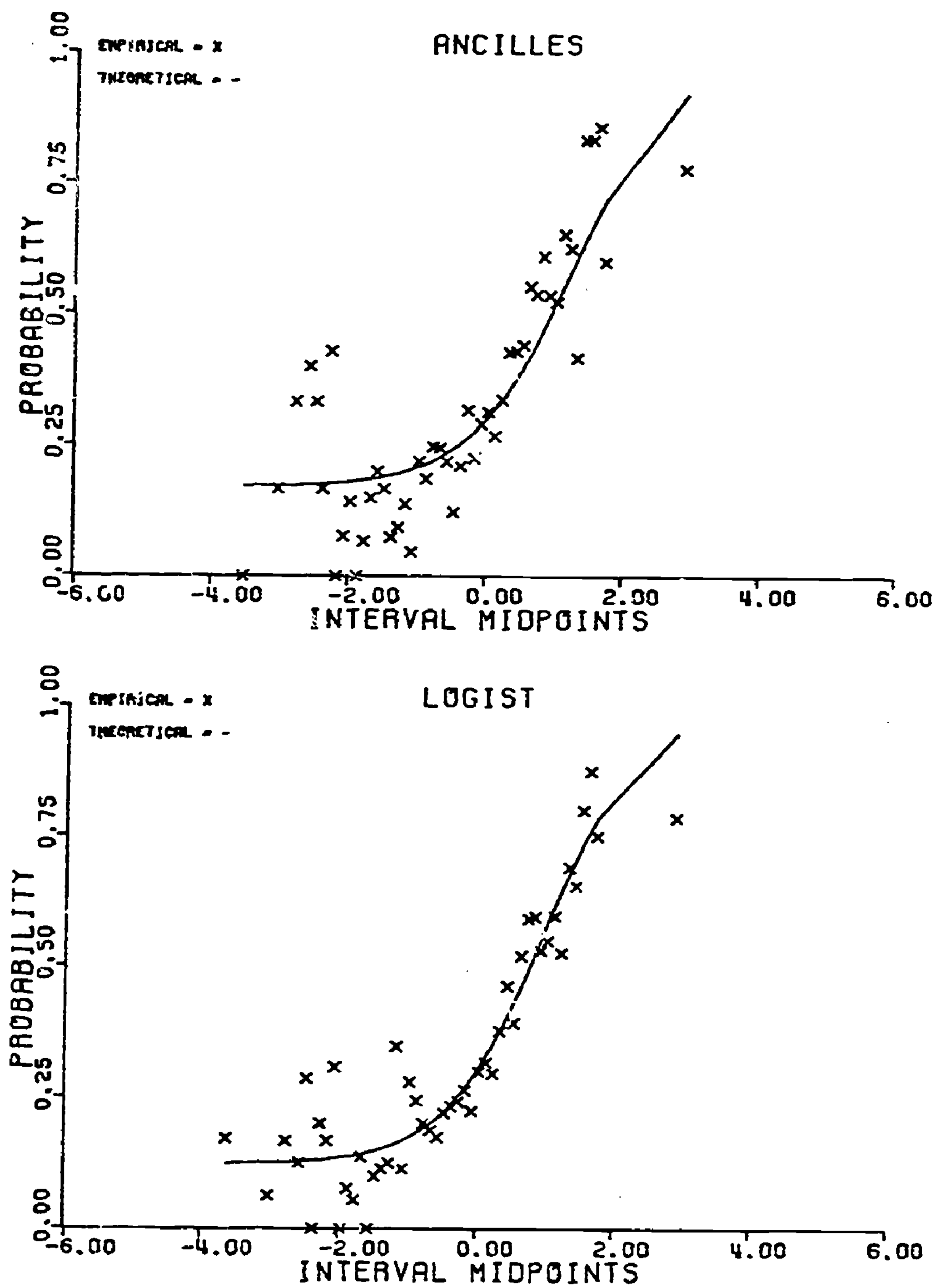


FIGURE 2

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 3

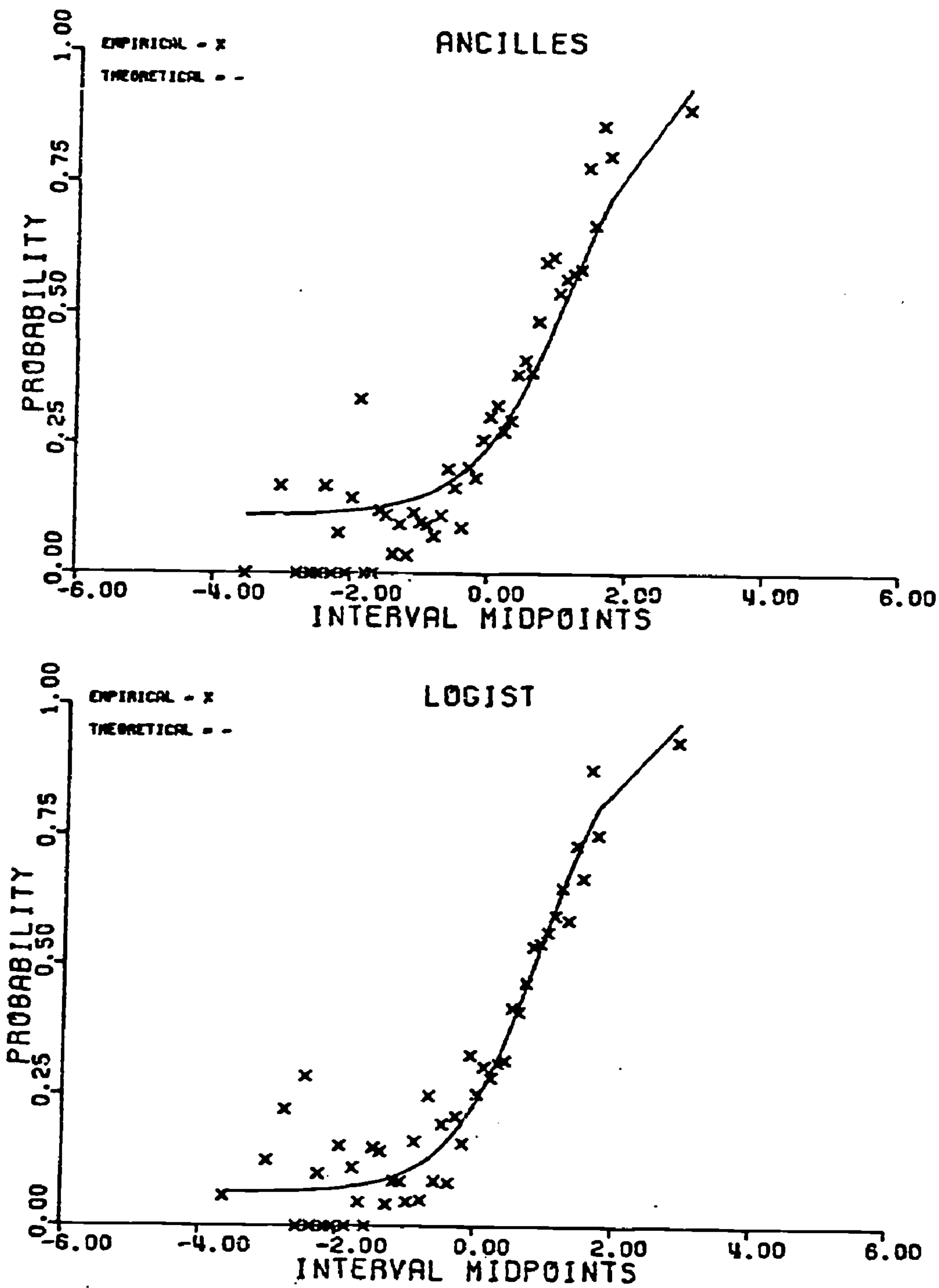


FIGURE 3

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 4

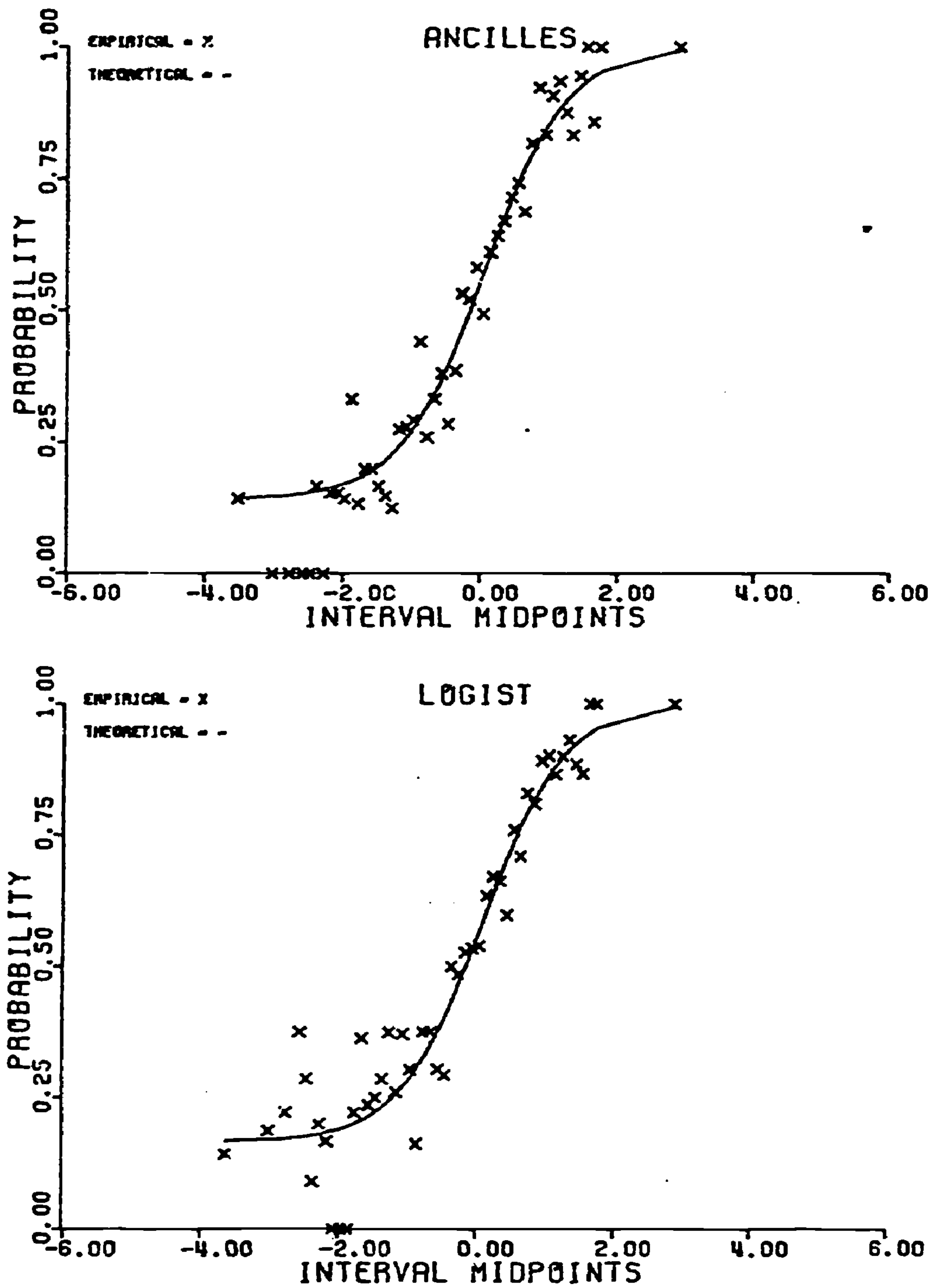


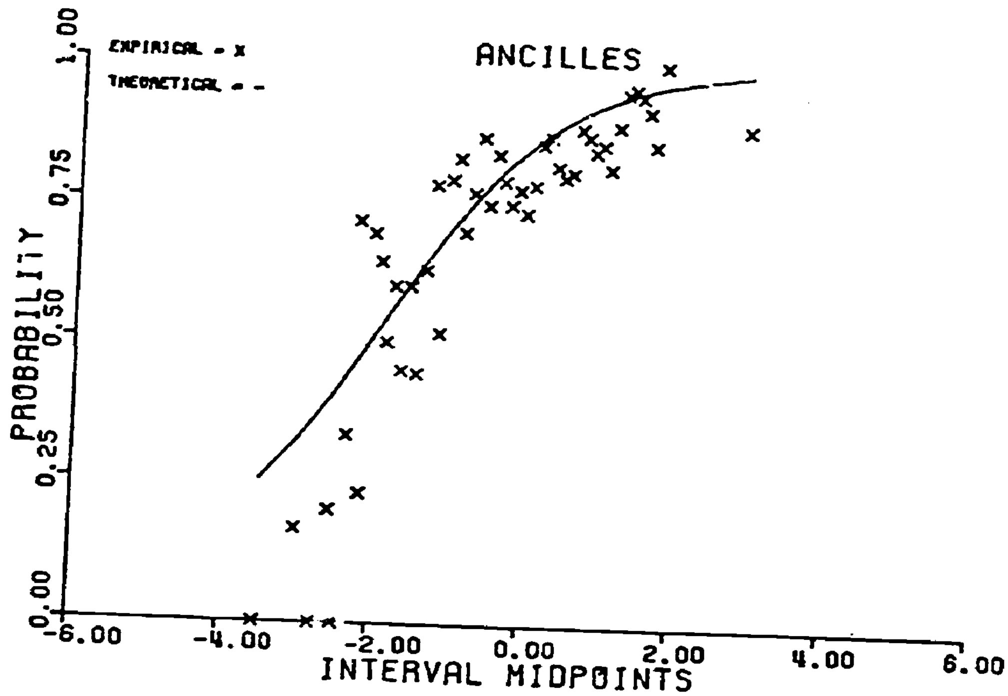
FIGURE 4

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 5



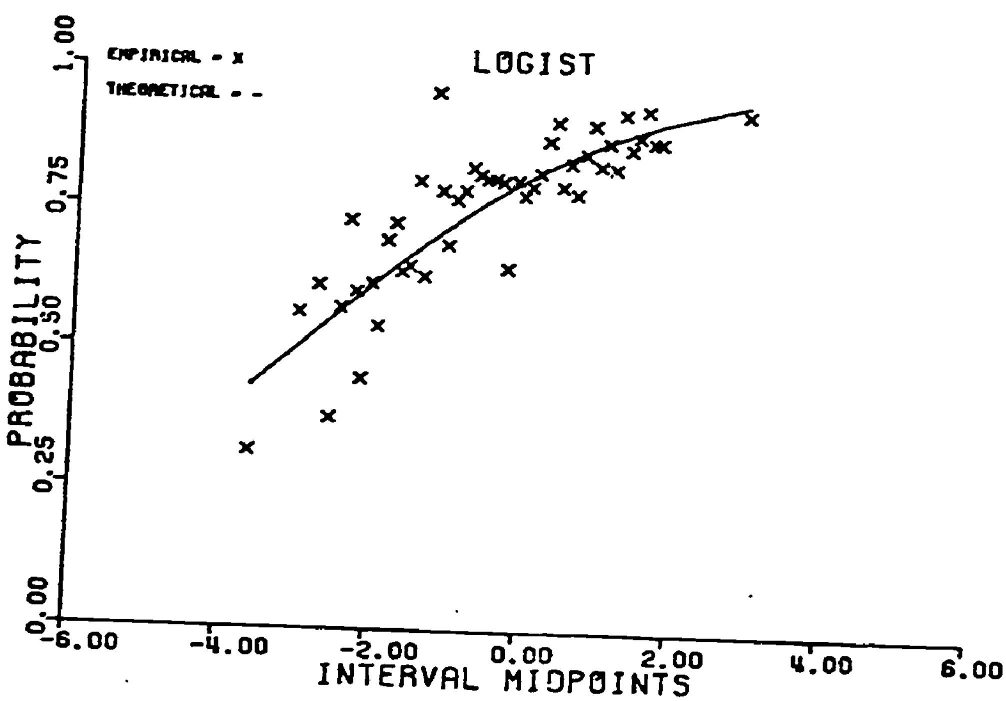


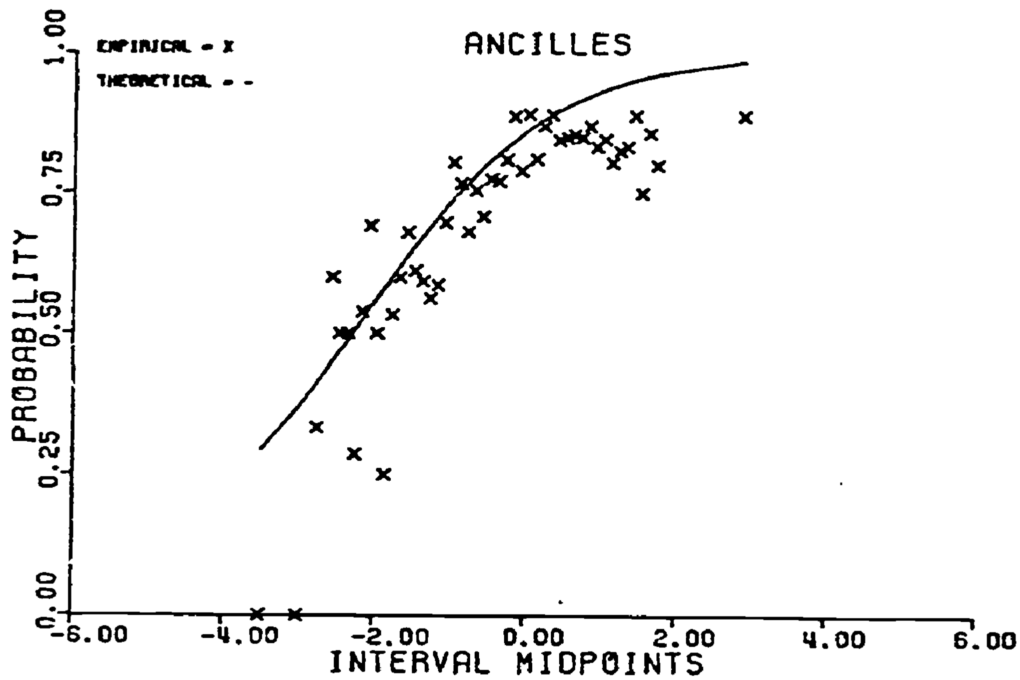
FIGURE 5

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 6



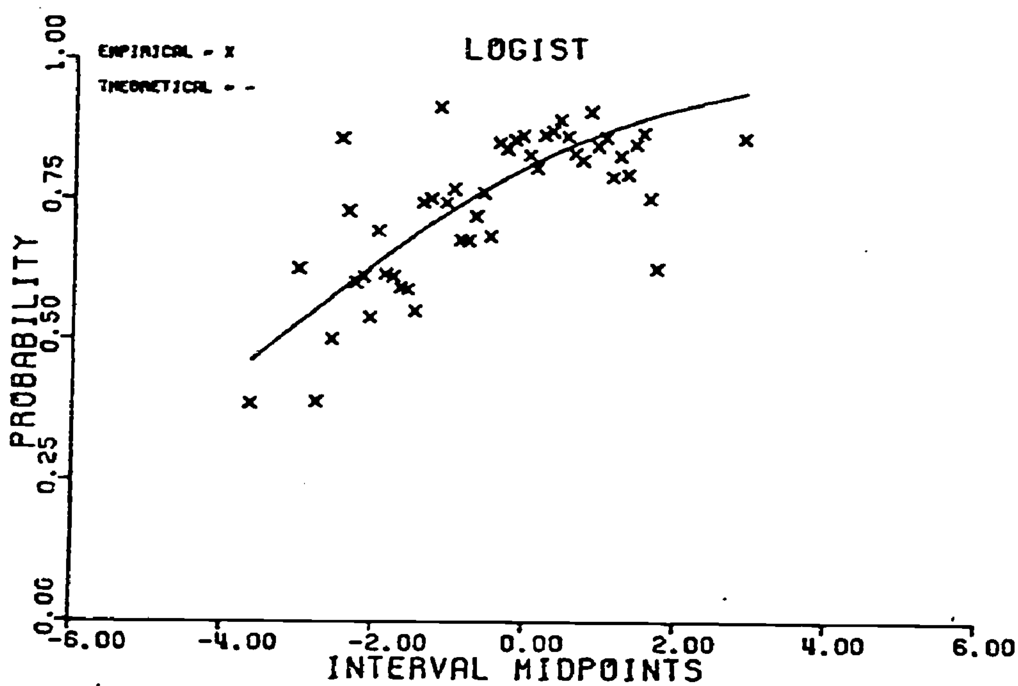


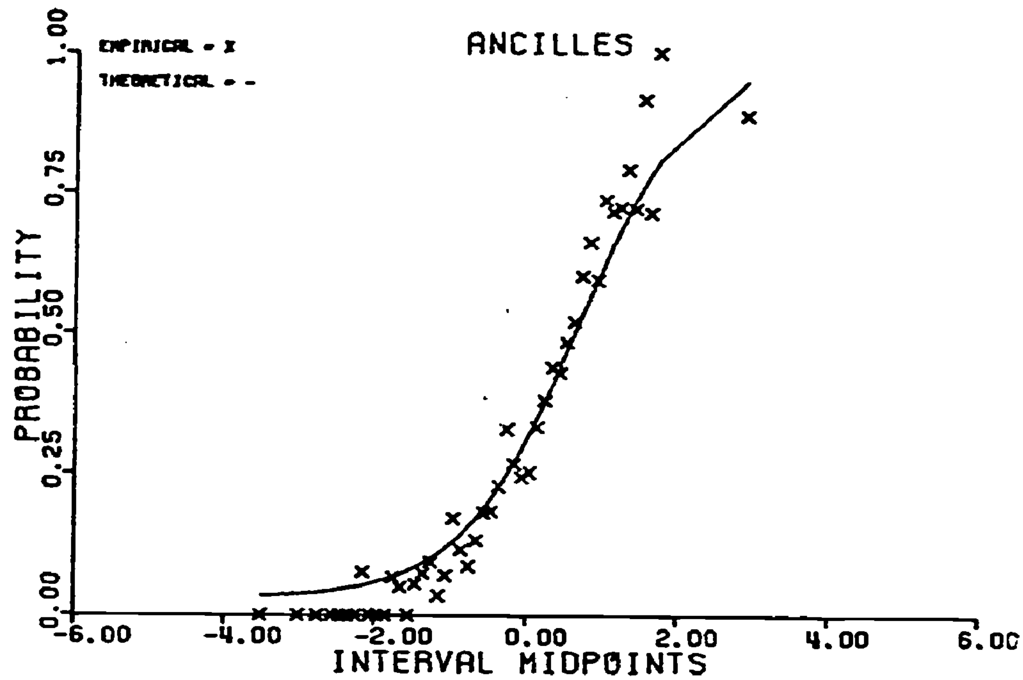
FIGURE 6

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 7



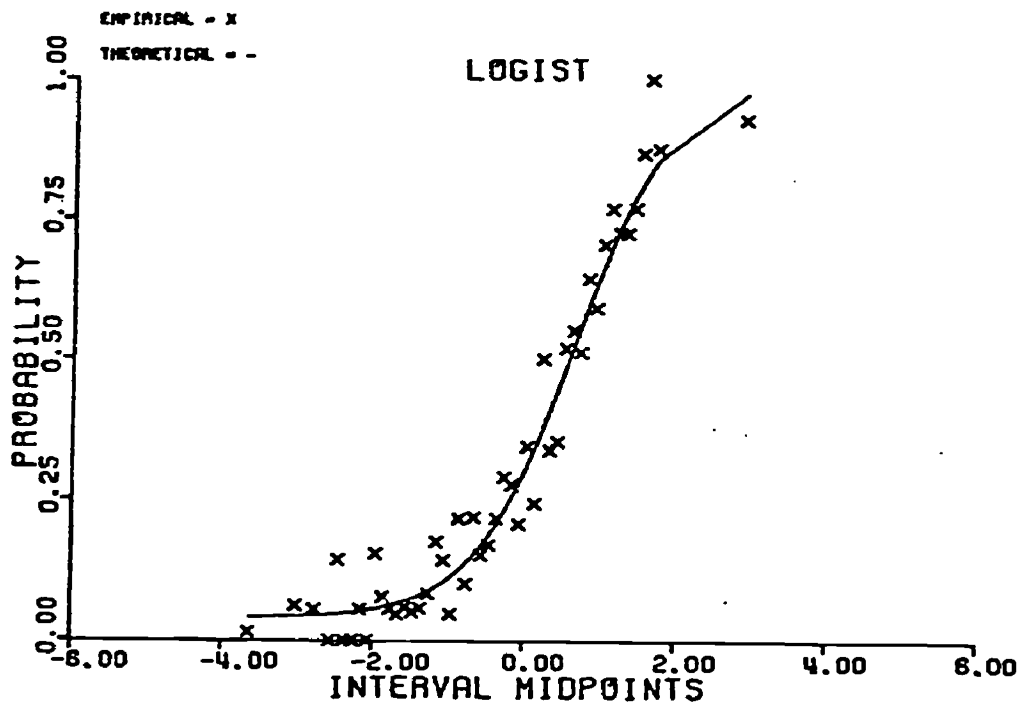


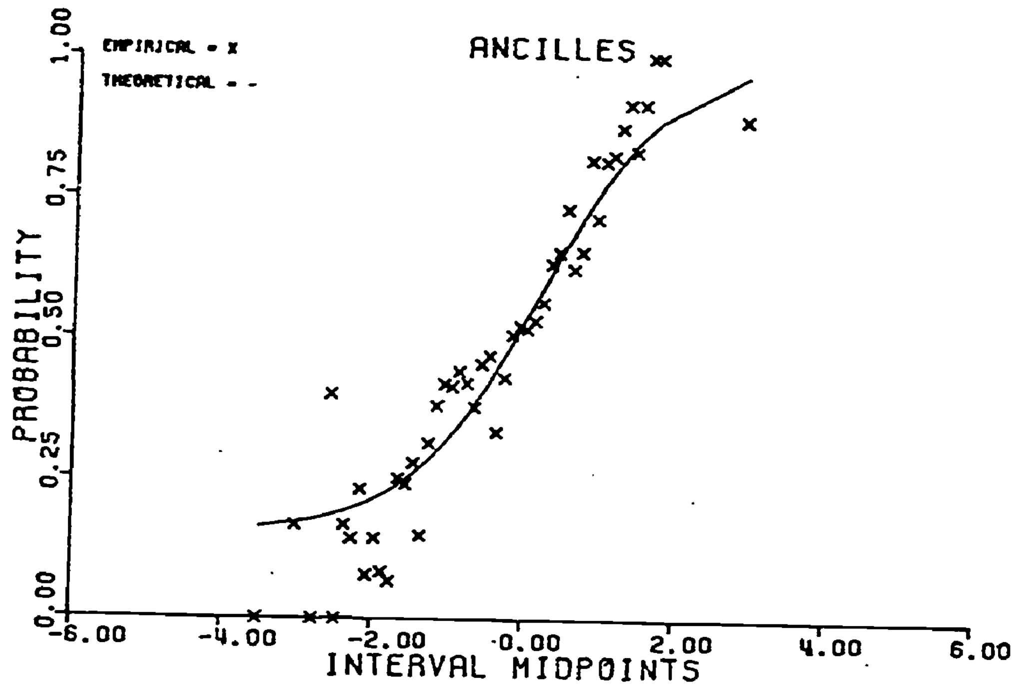
FIGURE 7

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 8



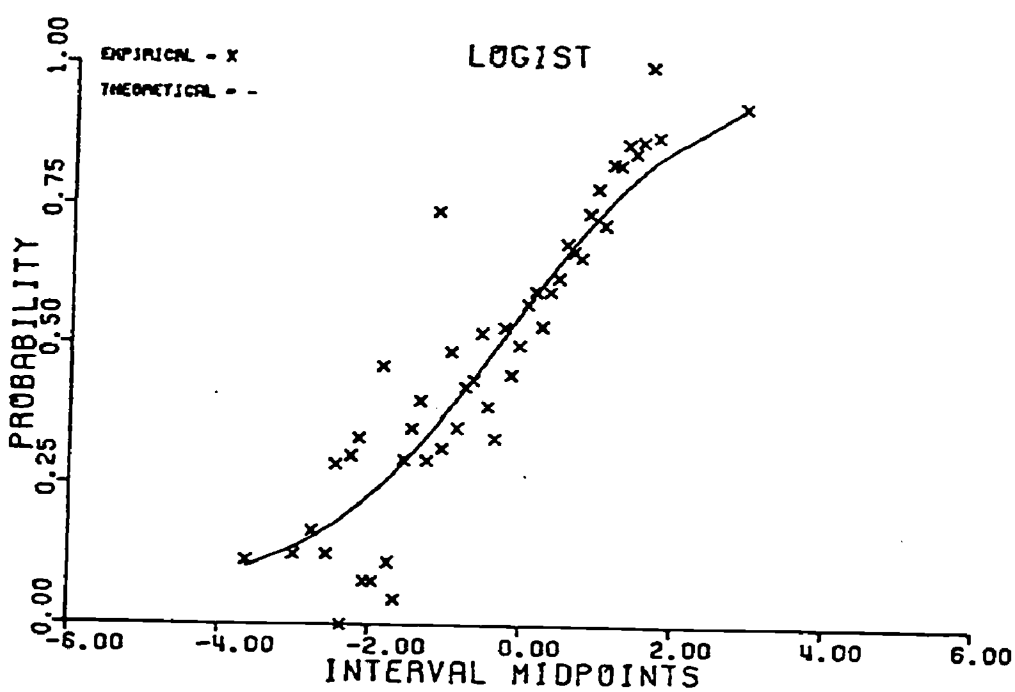


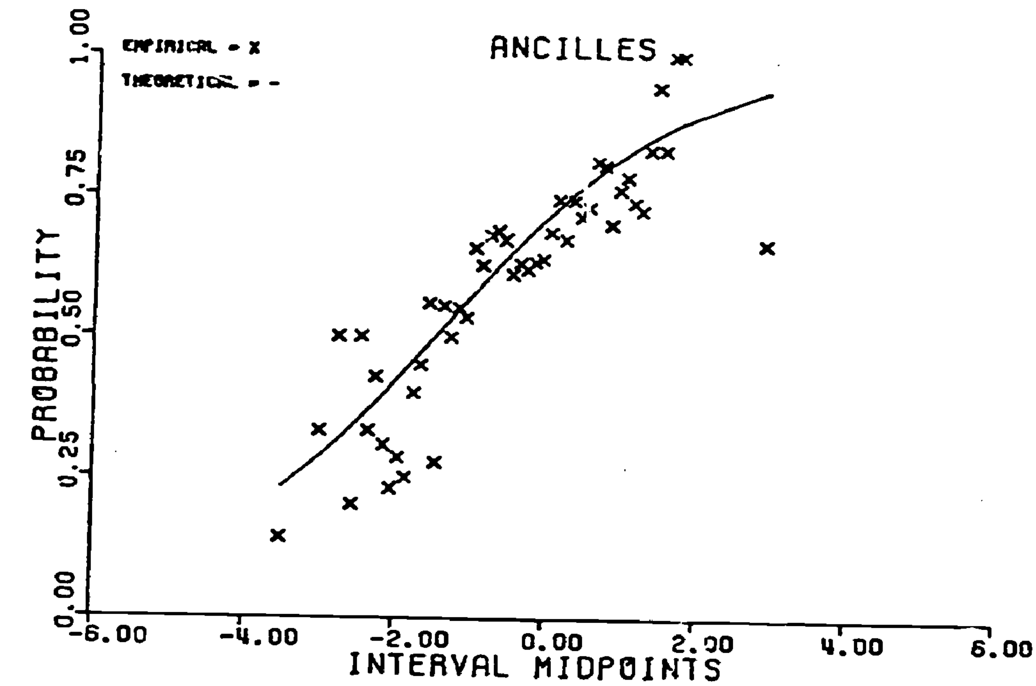
FIGURE 8

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 10



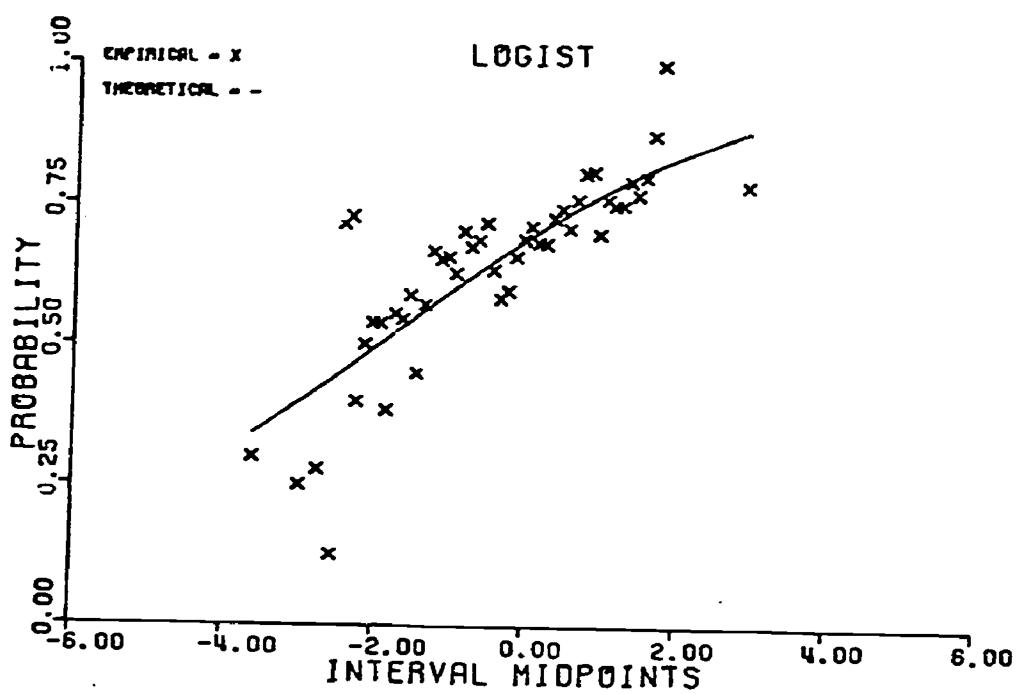


FIGURE 9

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 11

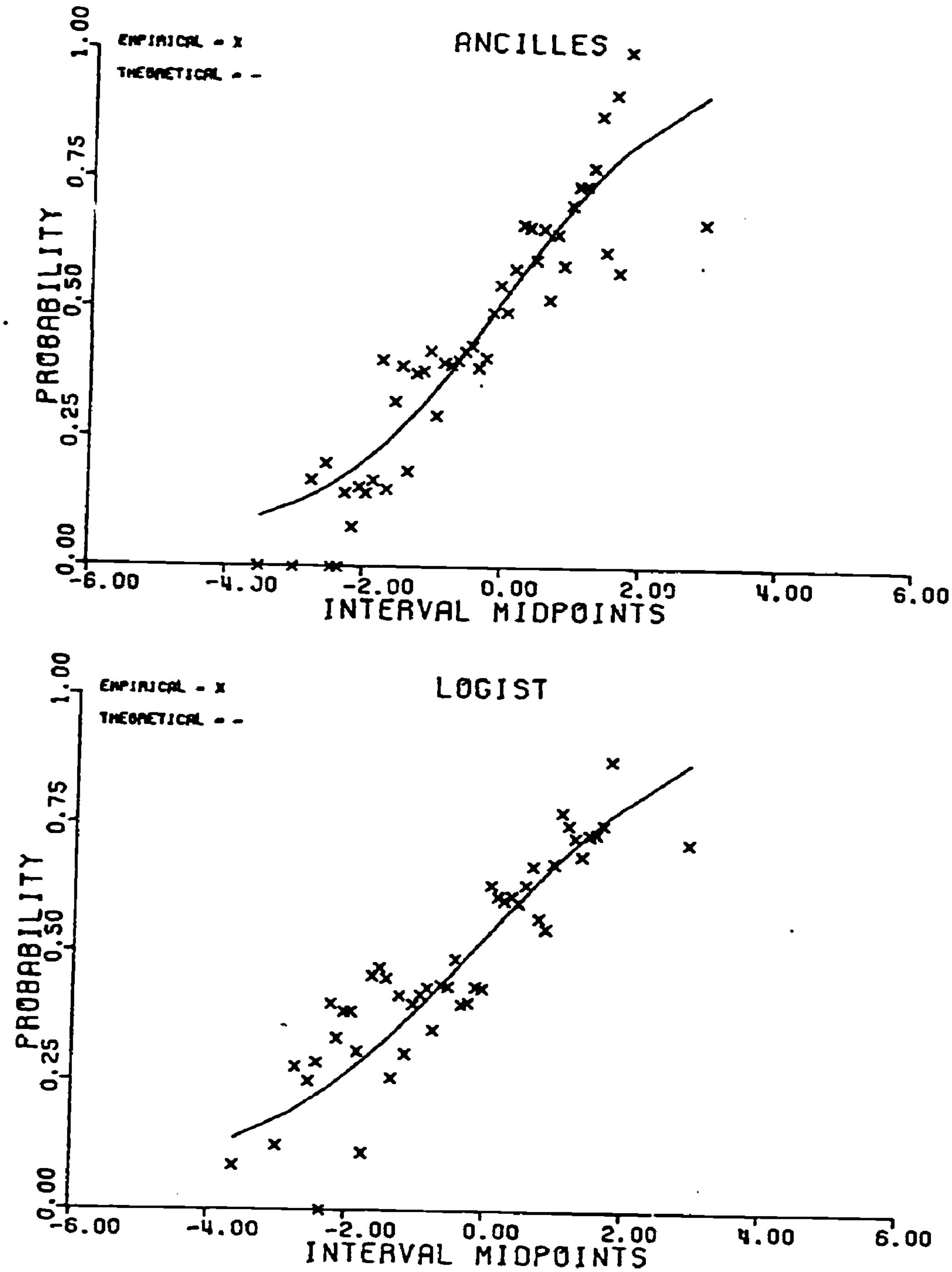
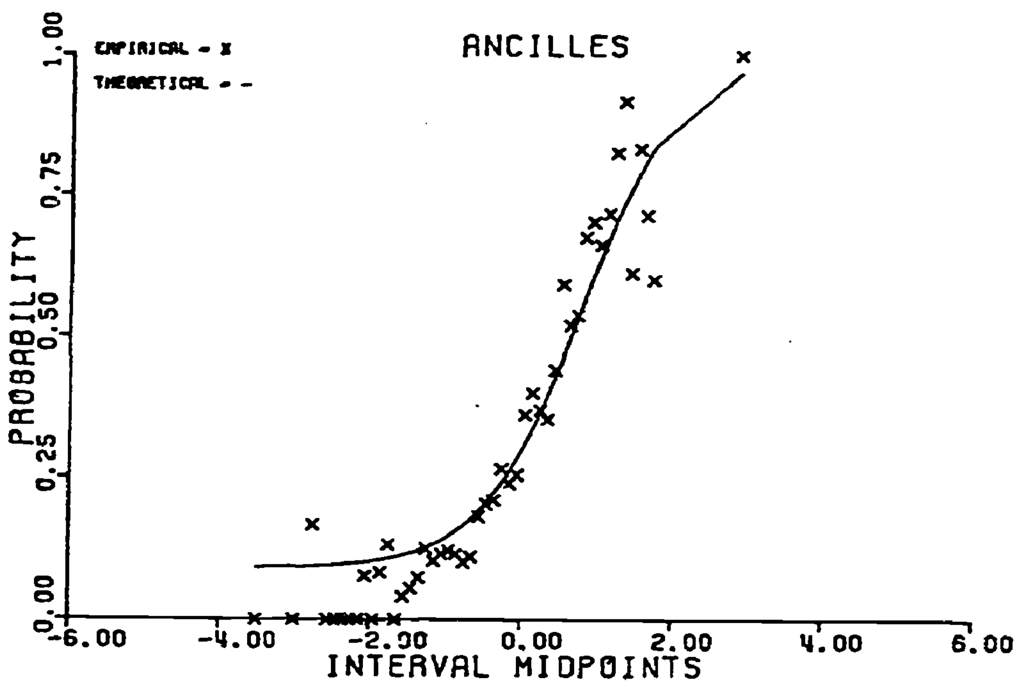


FIGURE 10

PLOTS OF EMPIRICAL AND THEORETICAL CURVES BASED ON ANCILLES AND LOGIST PROGRAMS

ITEM 12



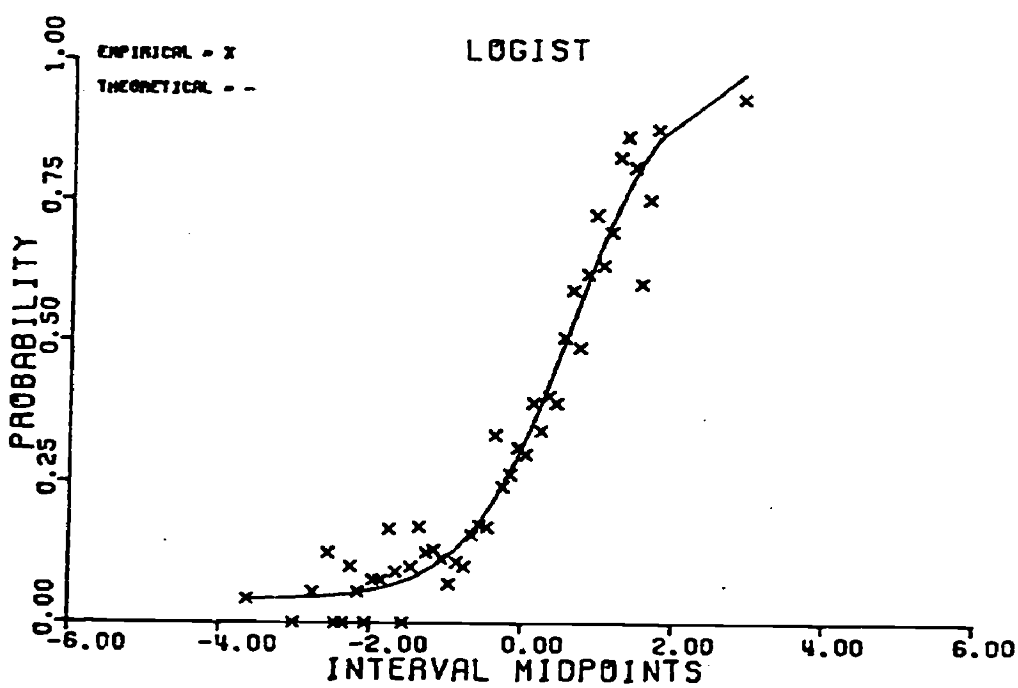


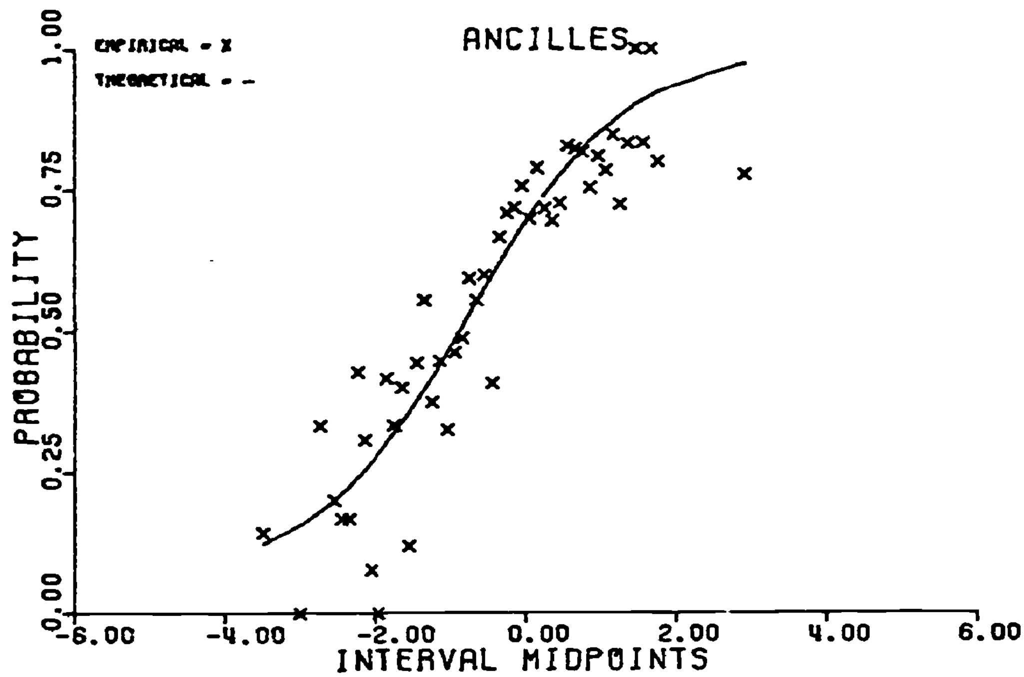
FIGURE 11

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 13



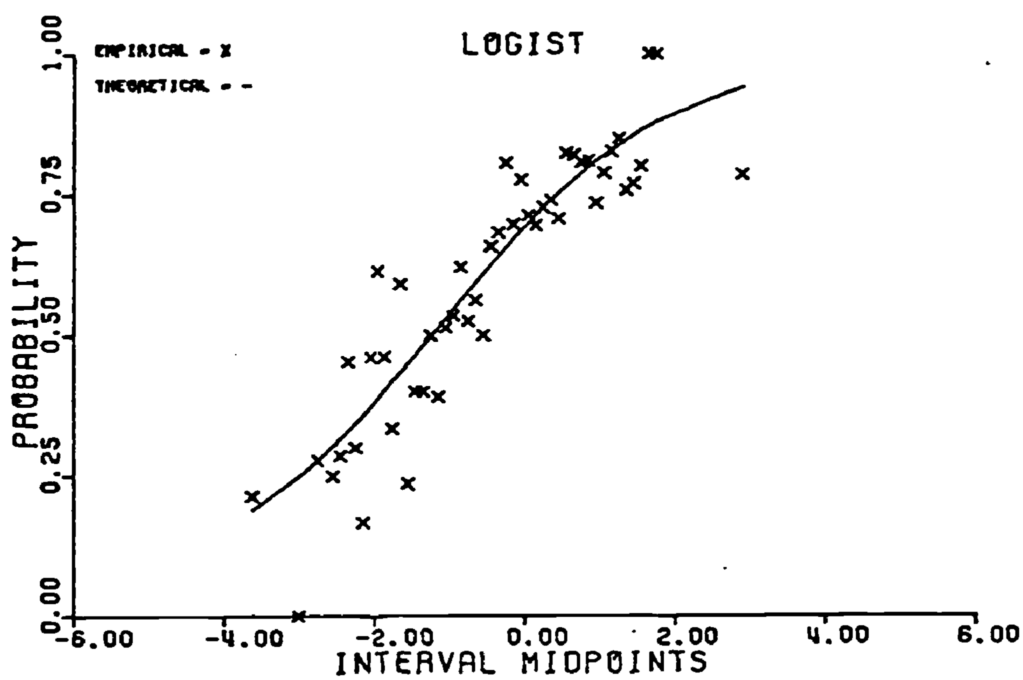


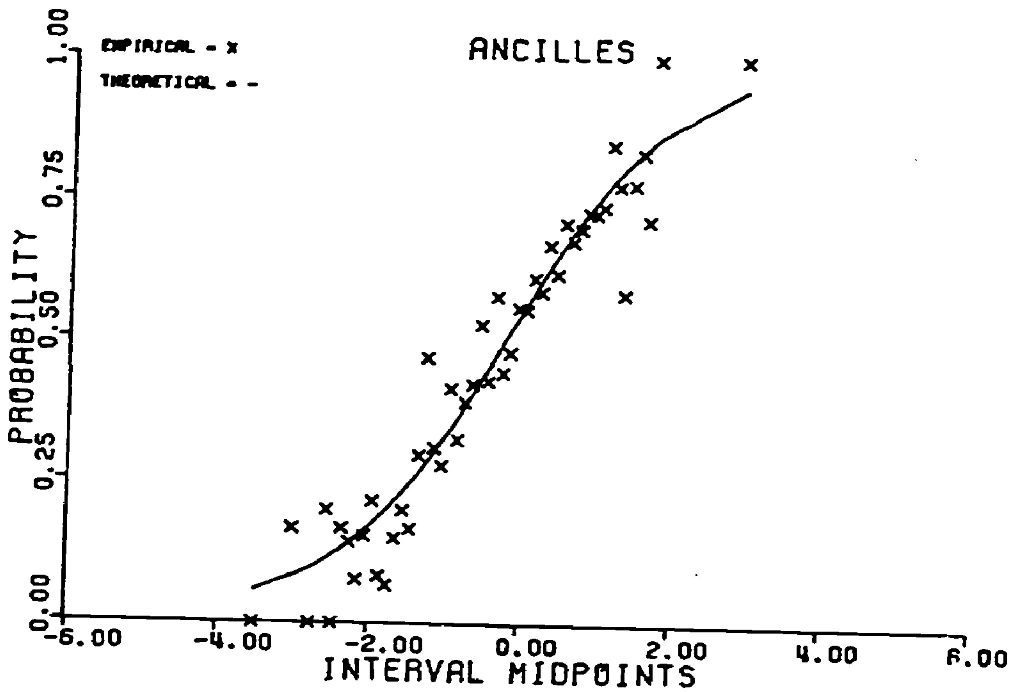
FIGURE 12

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 14



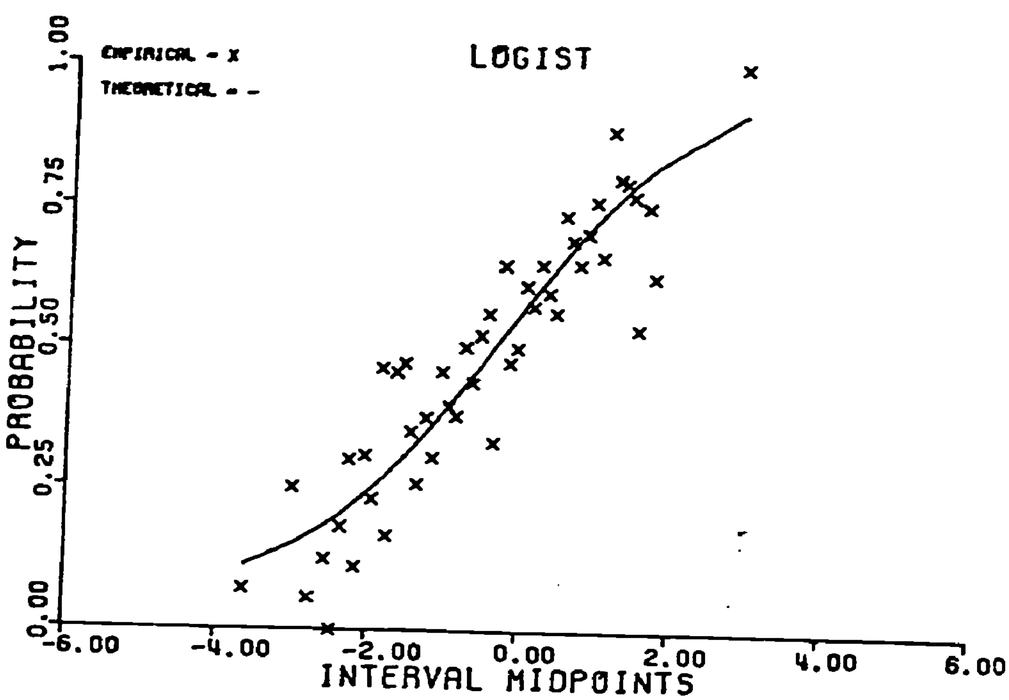


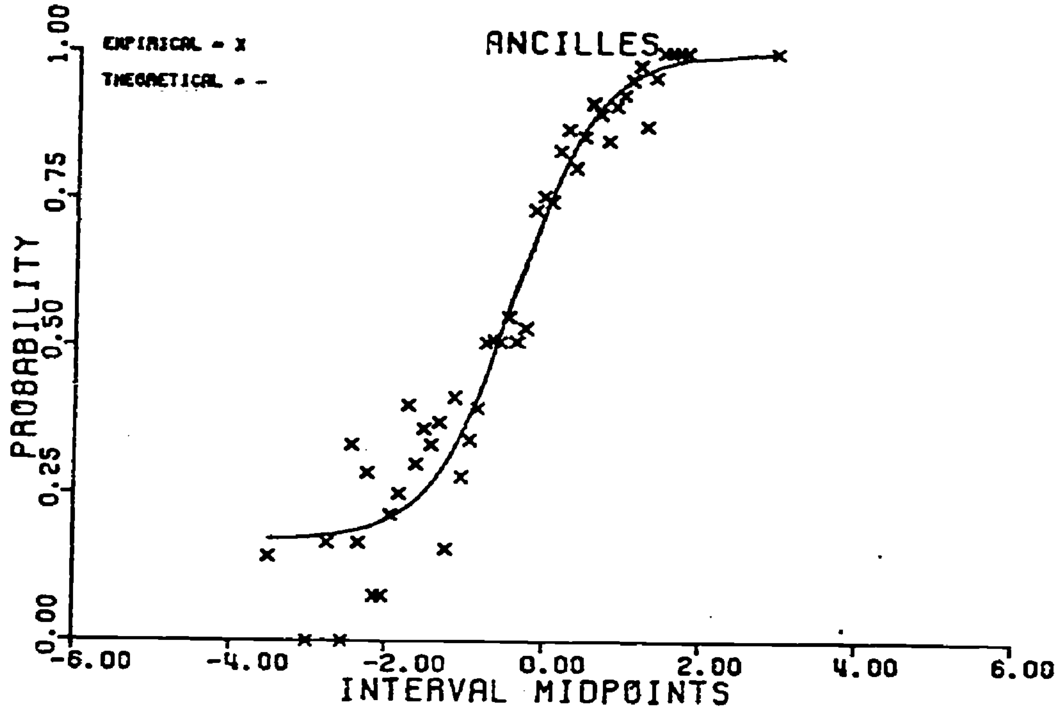
FIGURE 13

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 15



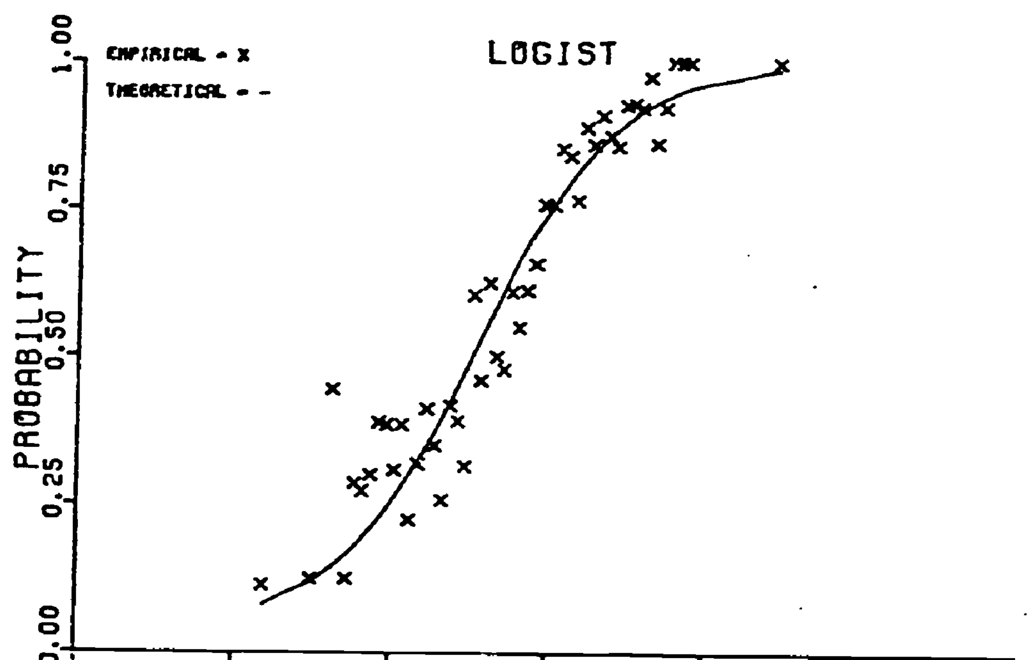


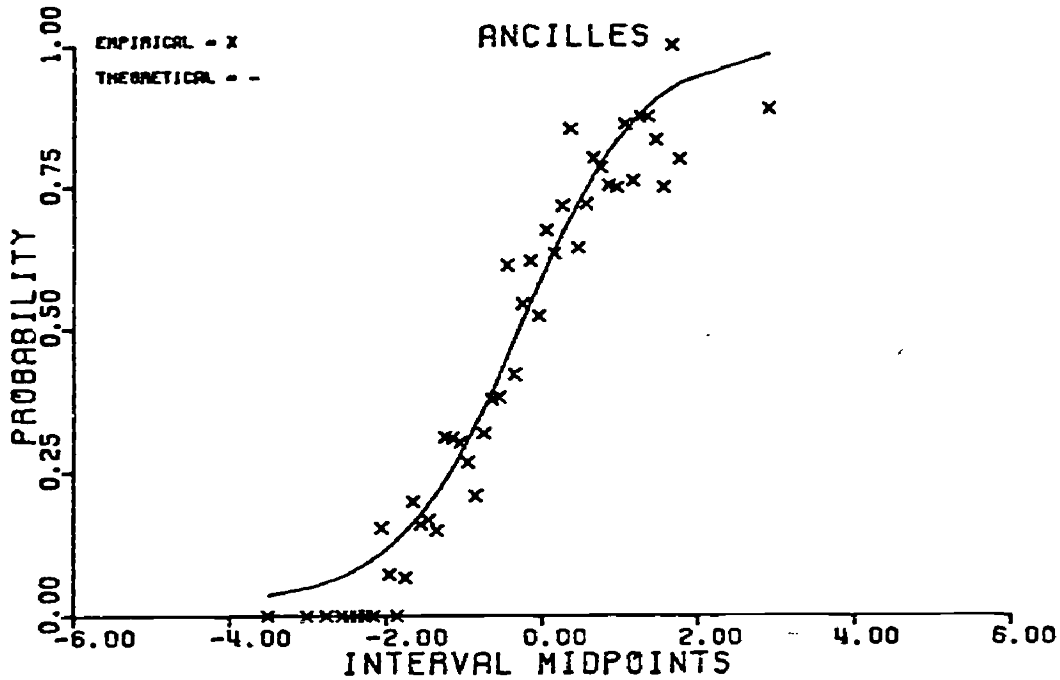
FIGURE 14

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 16



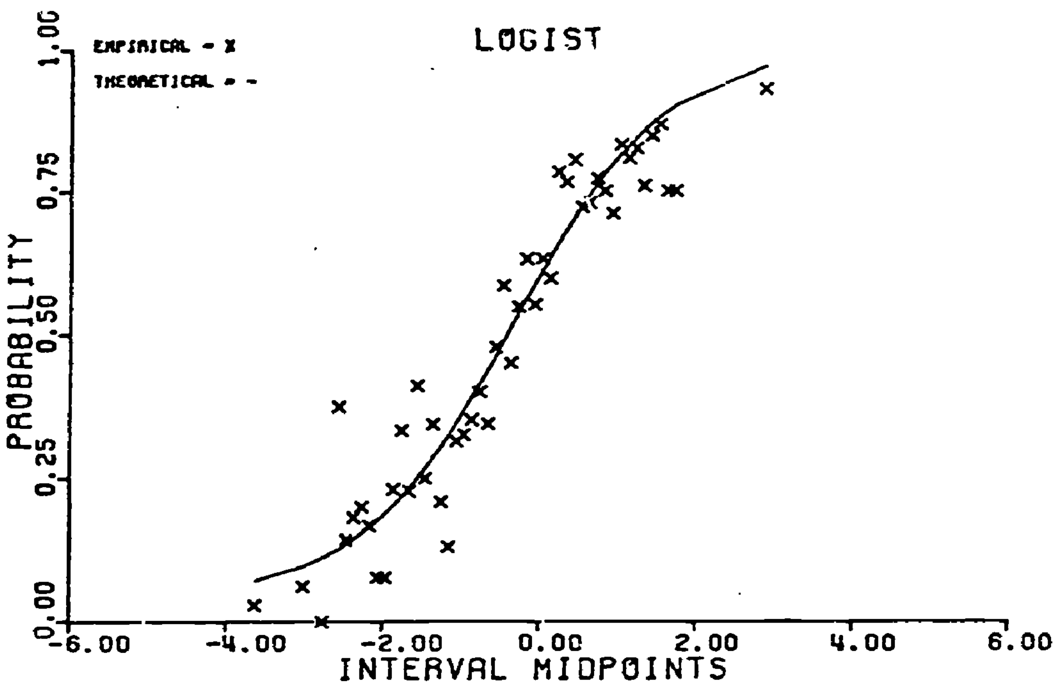


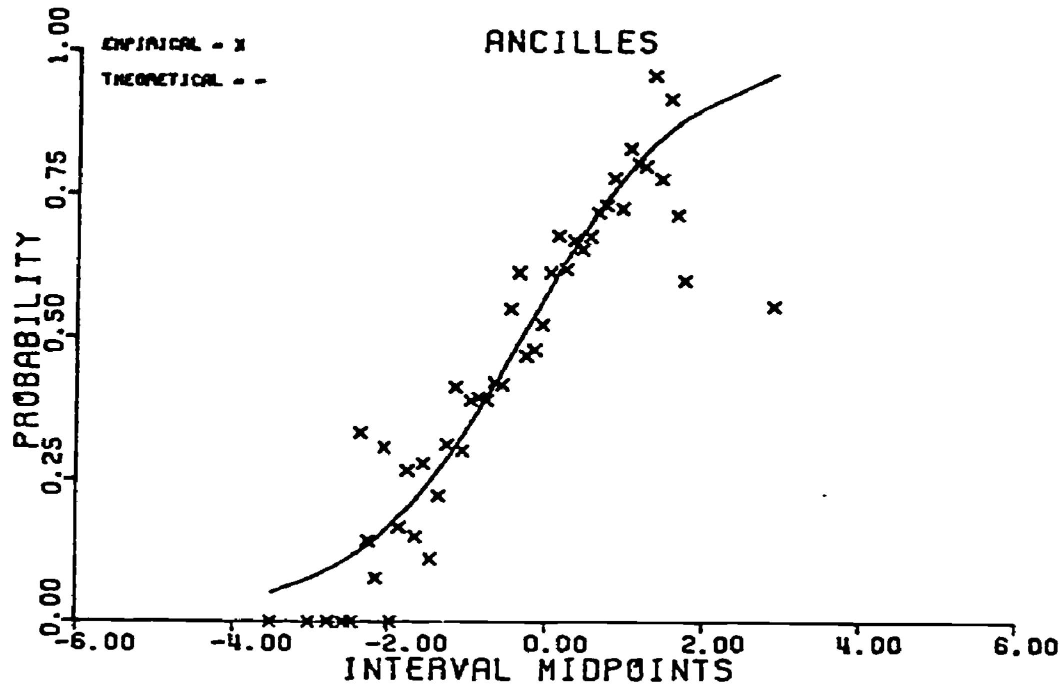
FIGURE 15

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 17



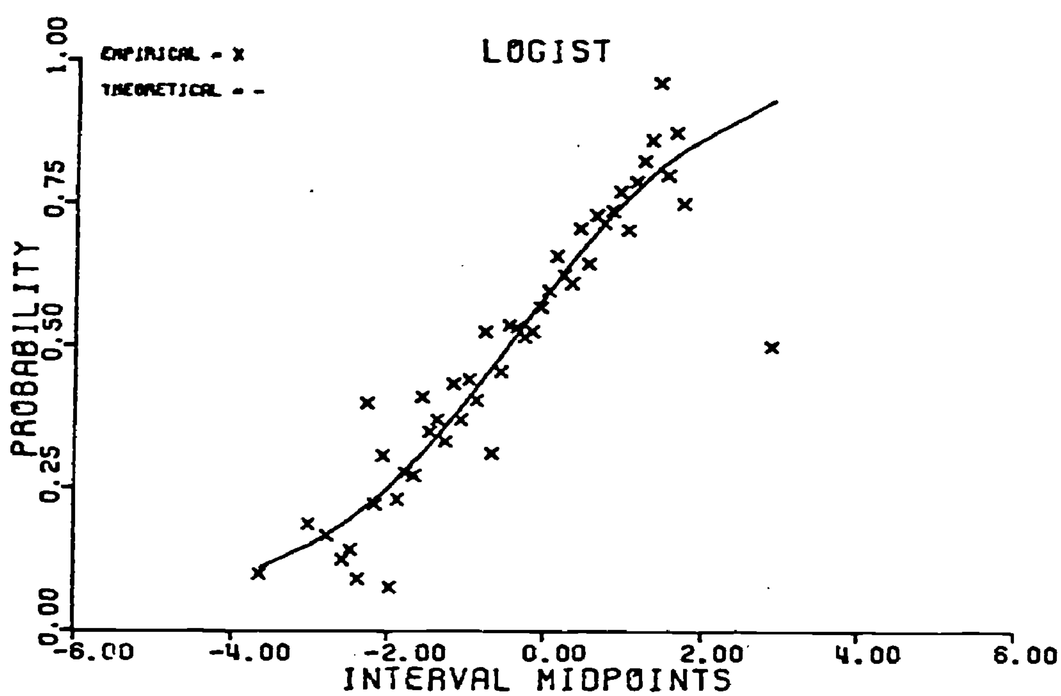


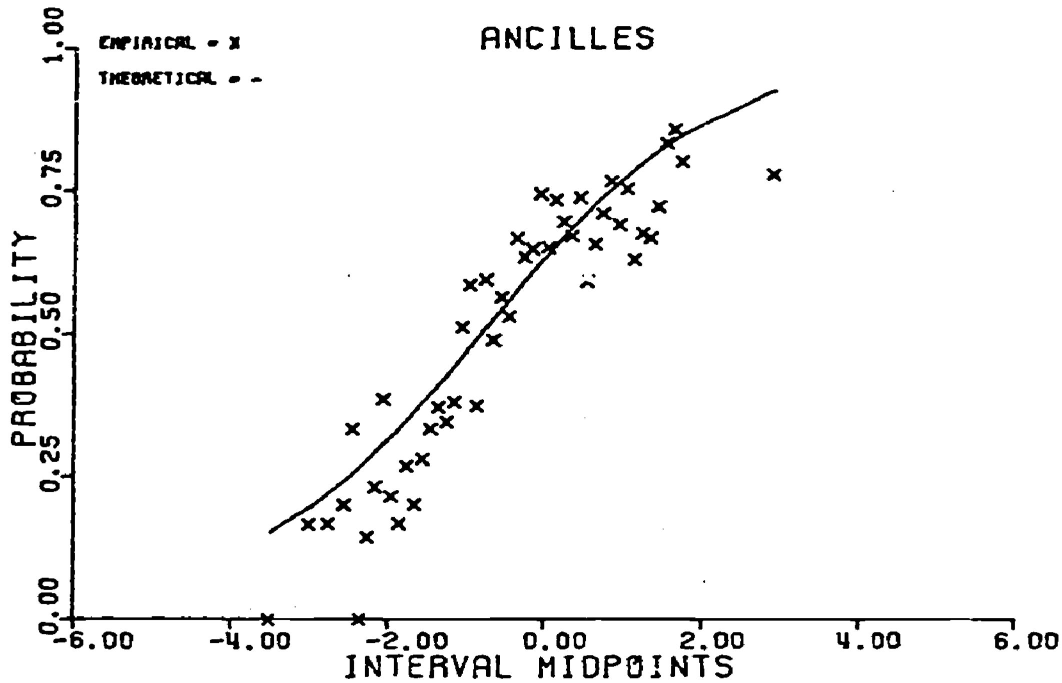
FIGURE 16

PLOTS OF EMPIRICAL AND

THEORETICAL ~~CURVES~~ BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 18



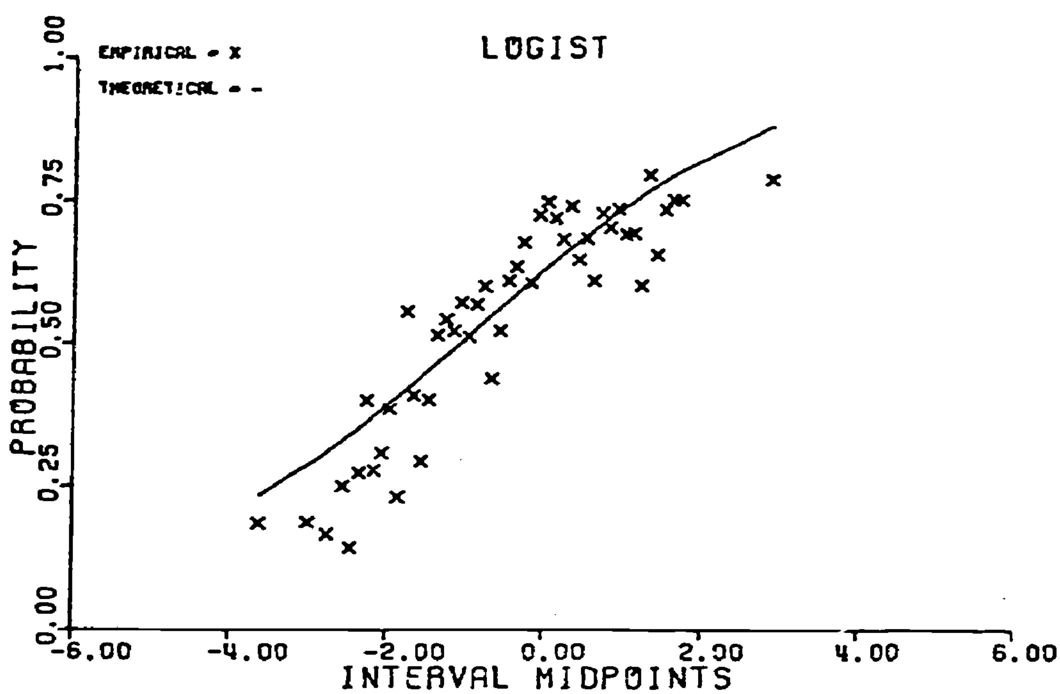


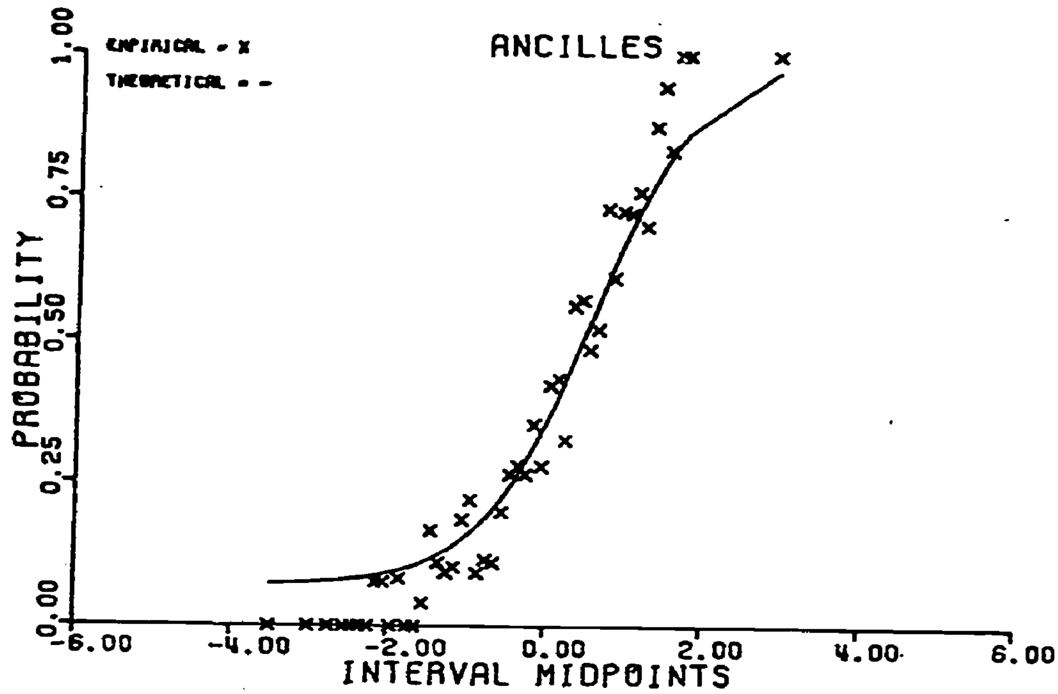
FIGURE 17

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 19



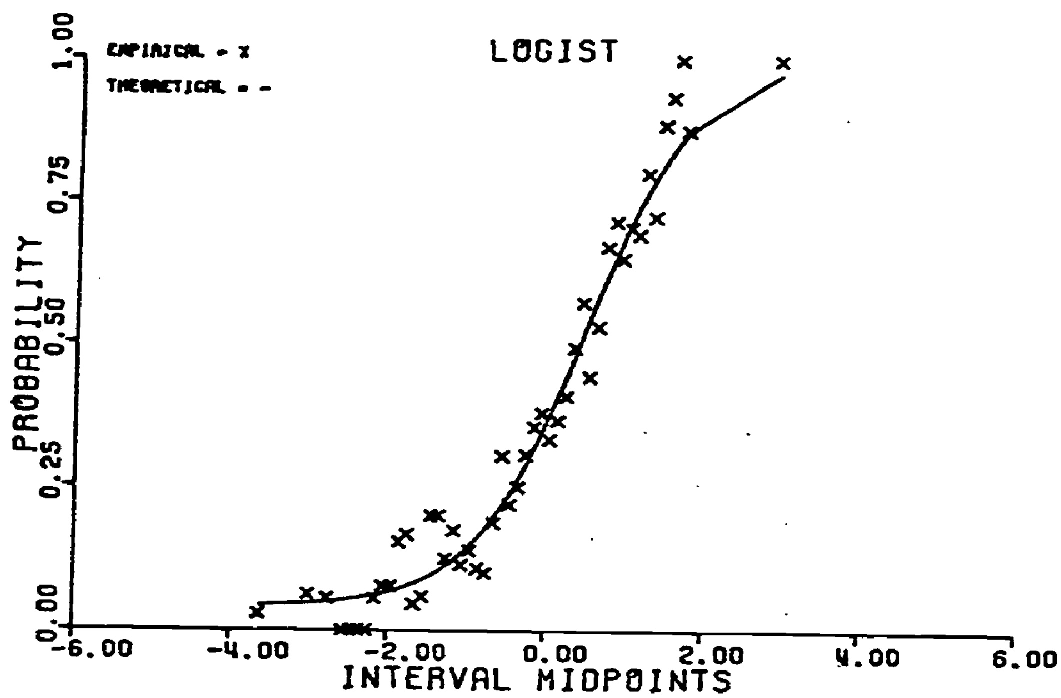
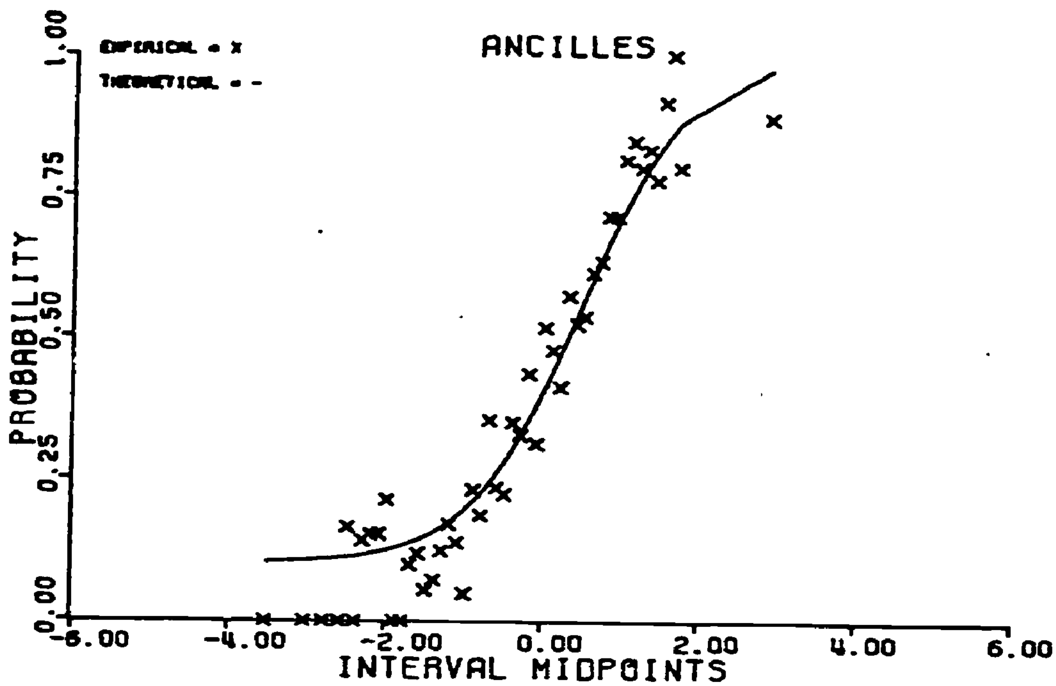


FIGURE 18

PLOTS OF EMPIRICAL AND THEORETICAL CURVES BASED ON ANCILLES AND LOGIST PROGRAMS

ITEM 20



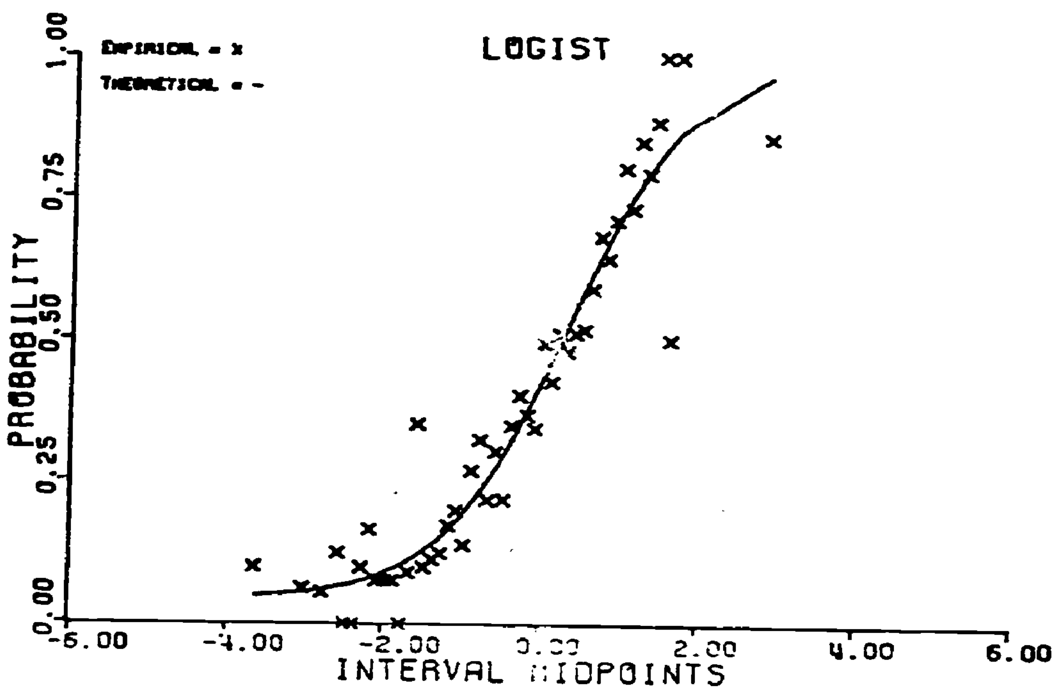


FIGURE 19

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 21

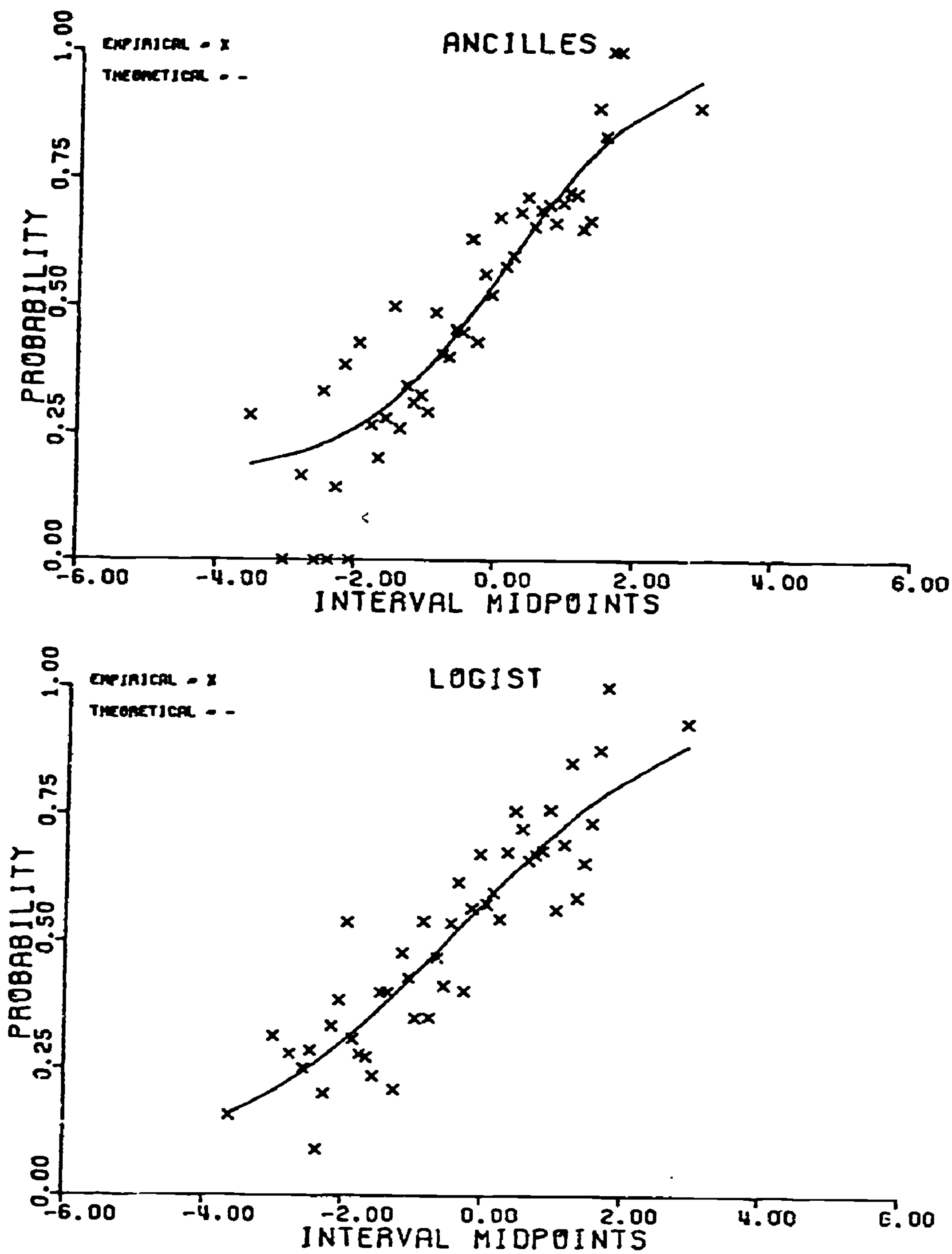
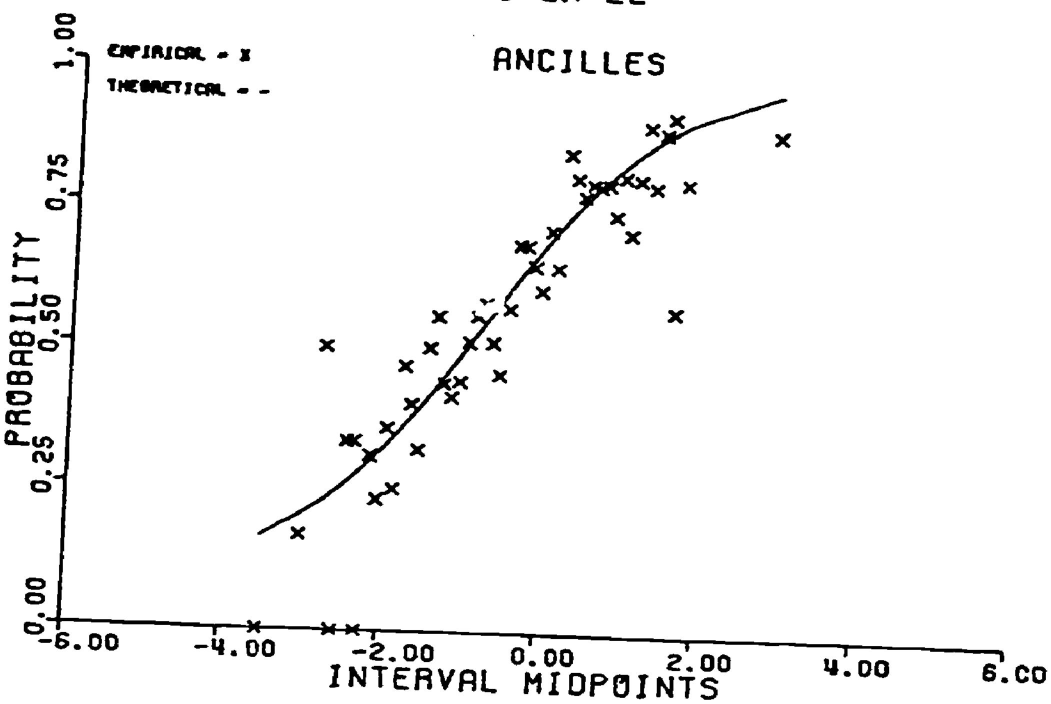


FIGURE 20

PLOTS OF EMPIRICAL AND THEORETICAL CURVES BASED ON ANCILLES AND LOGIST PROGRAMS

ITEM 22



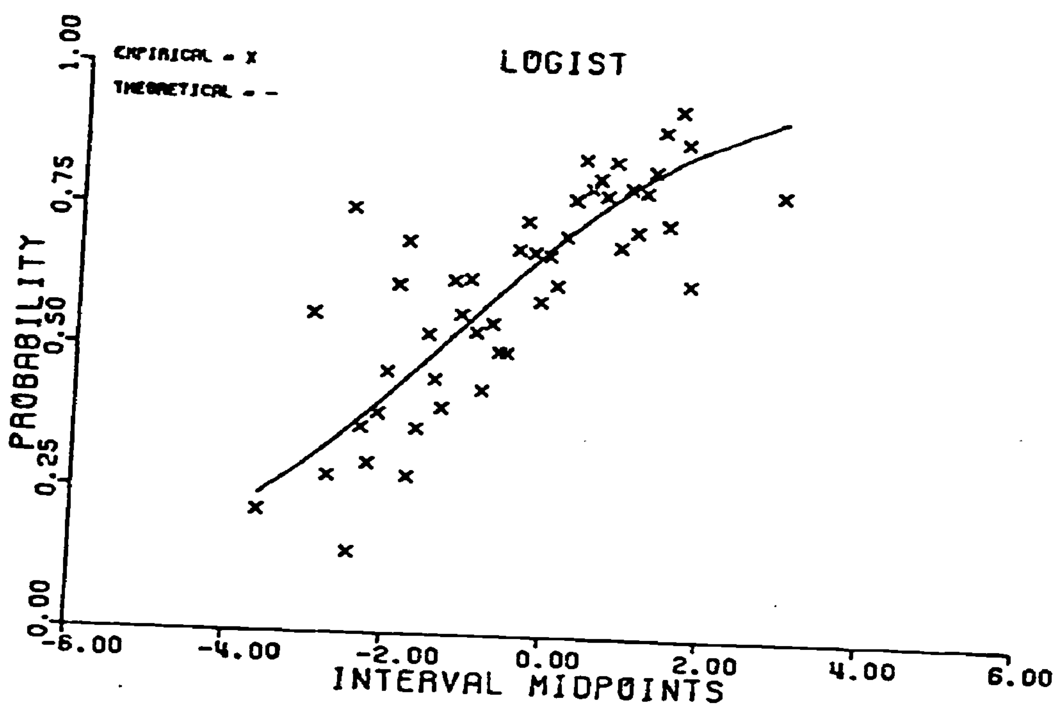


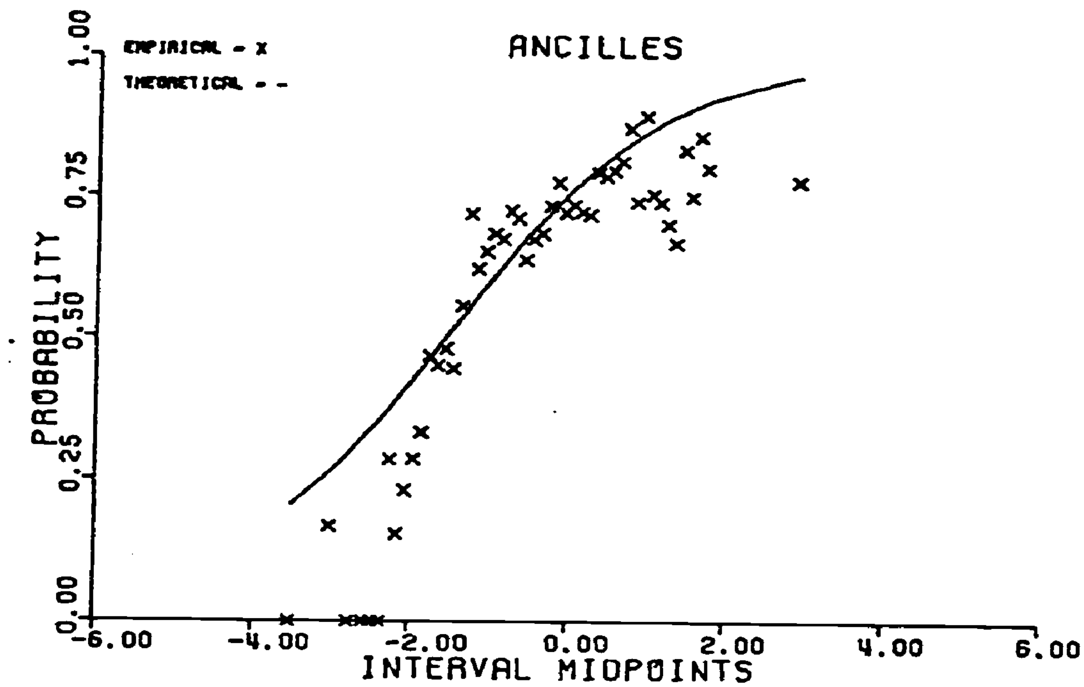
FIGURE 21

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 23



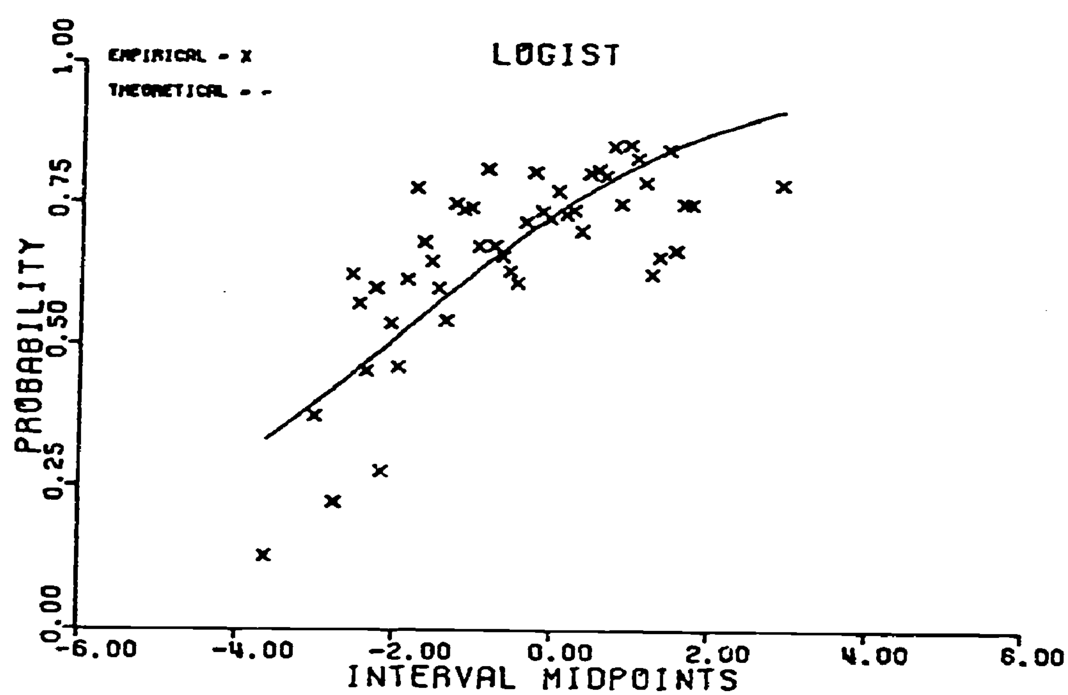


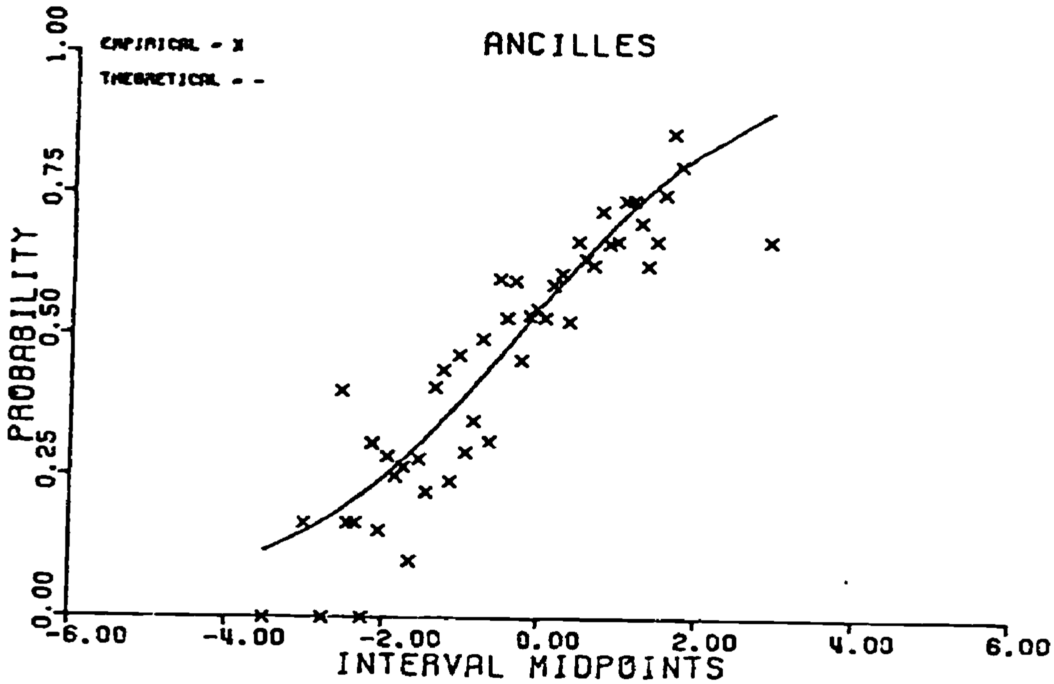
FIGURE 22

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 24



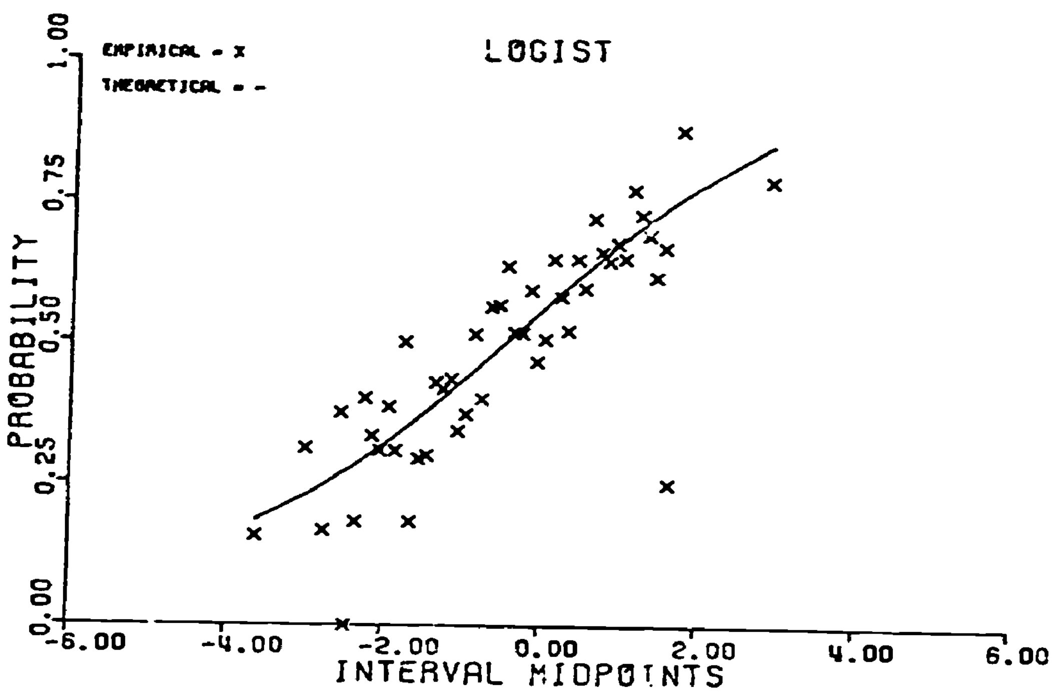


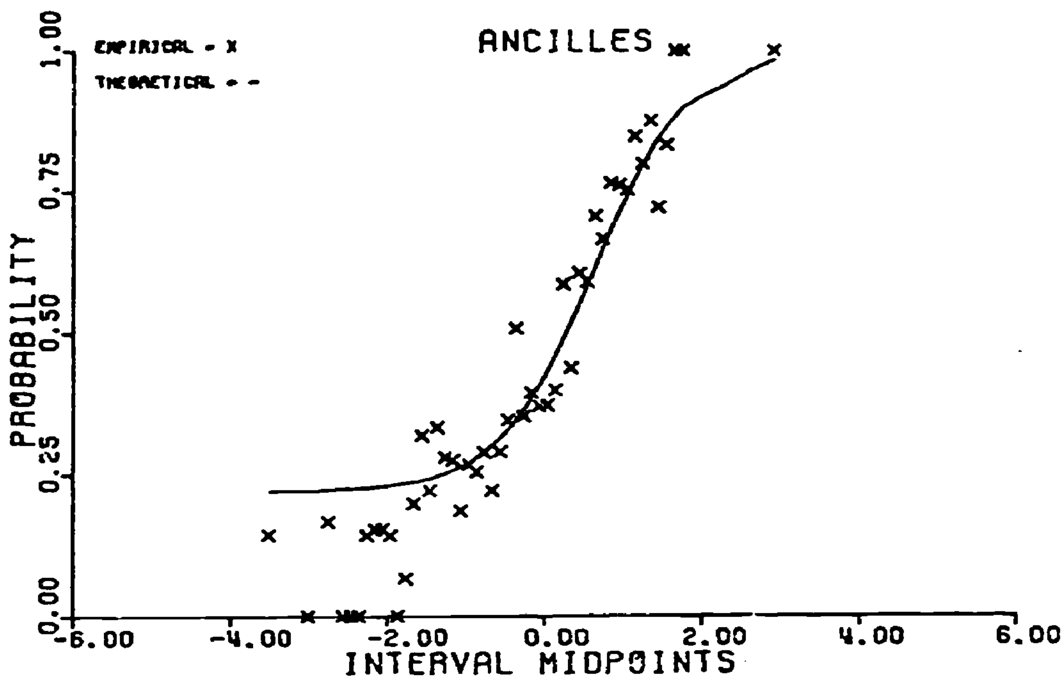
FIGURE 23

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLFS AND LOGIST PROGRAMS

ITEM 25



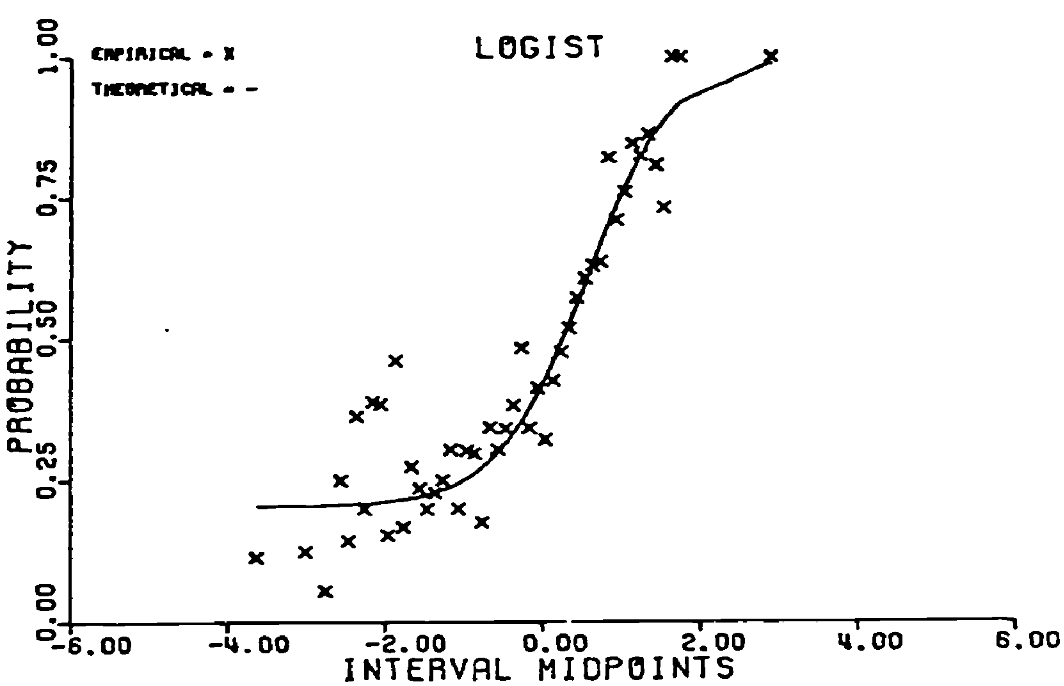


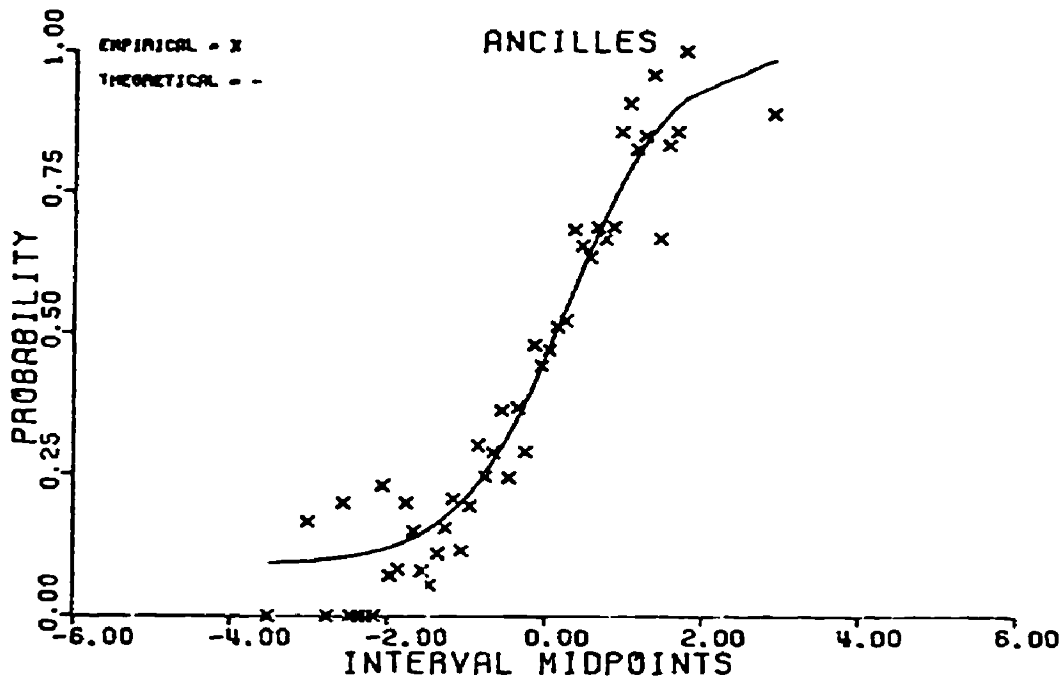
FIGURE 24

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 26



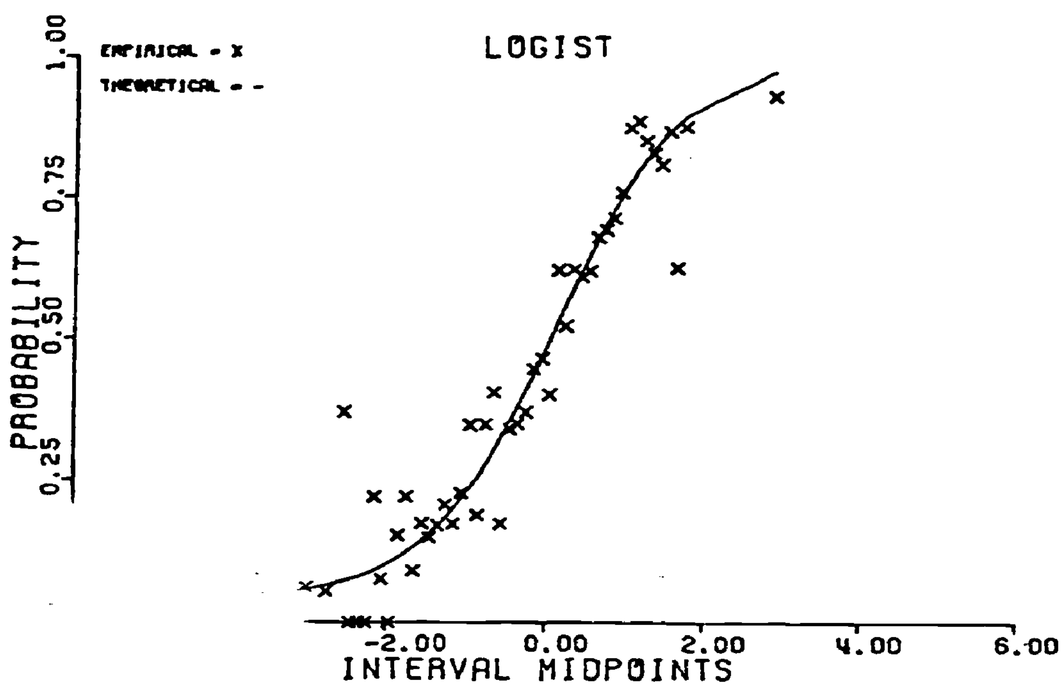


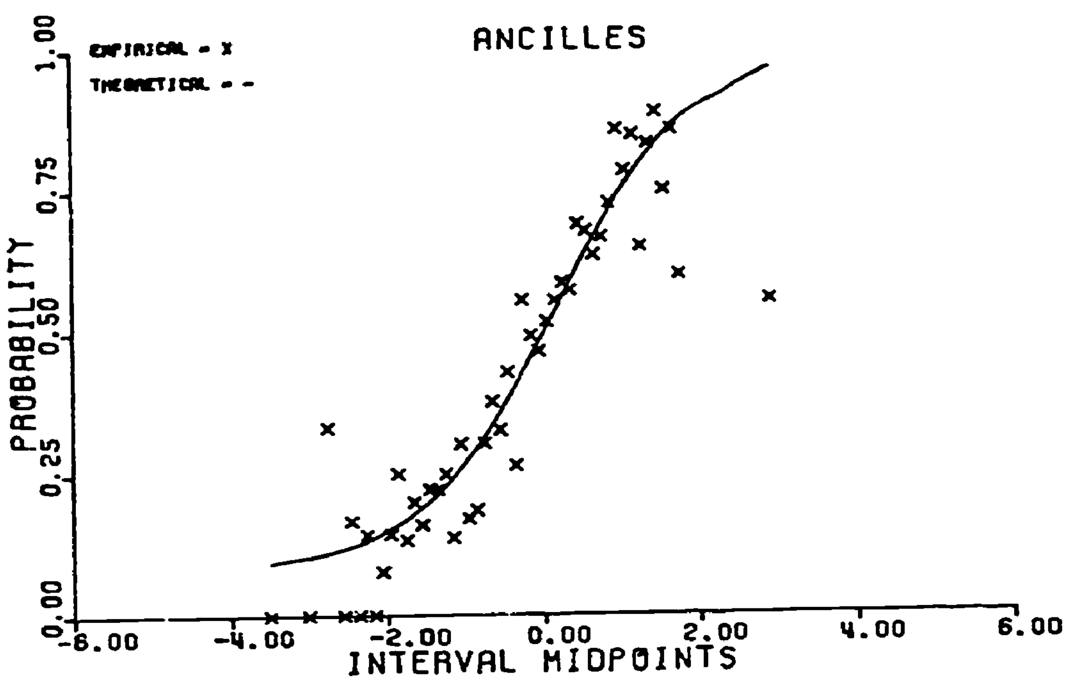
FIGURE 25

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 27



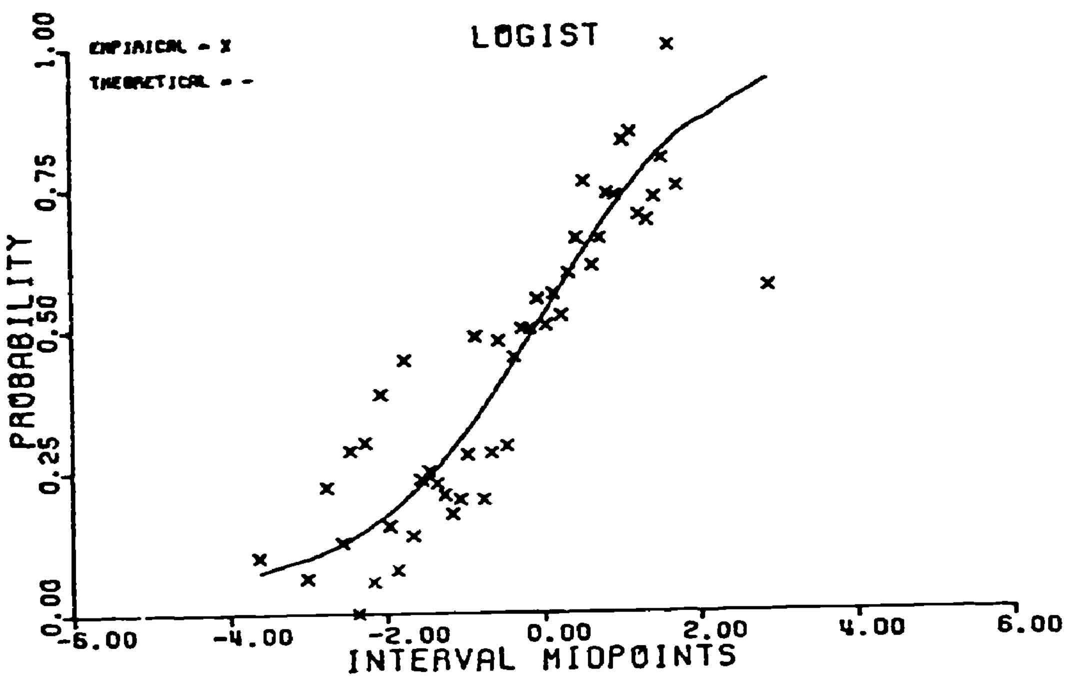


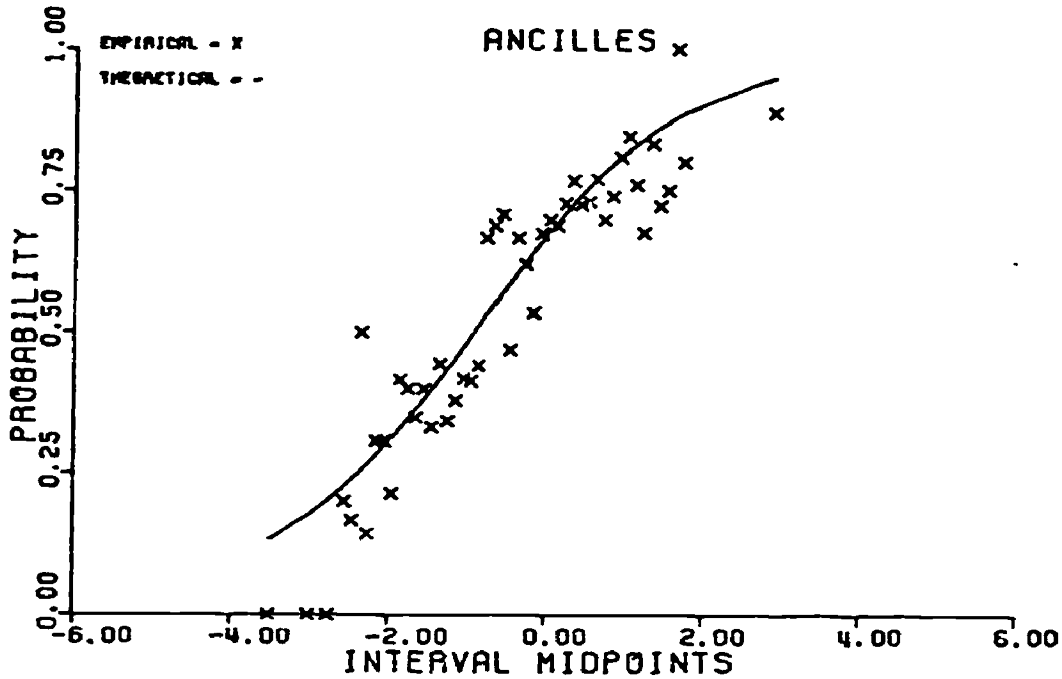
FIGURE 26

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 28



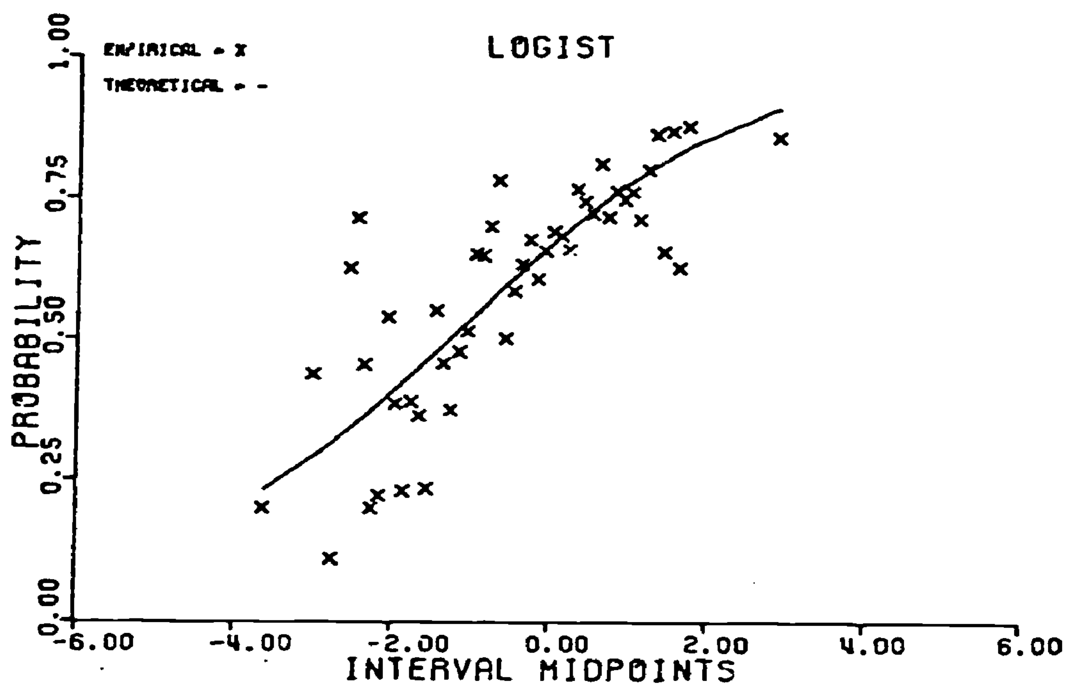


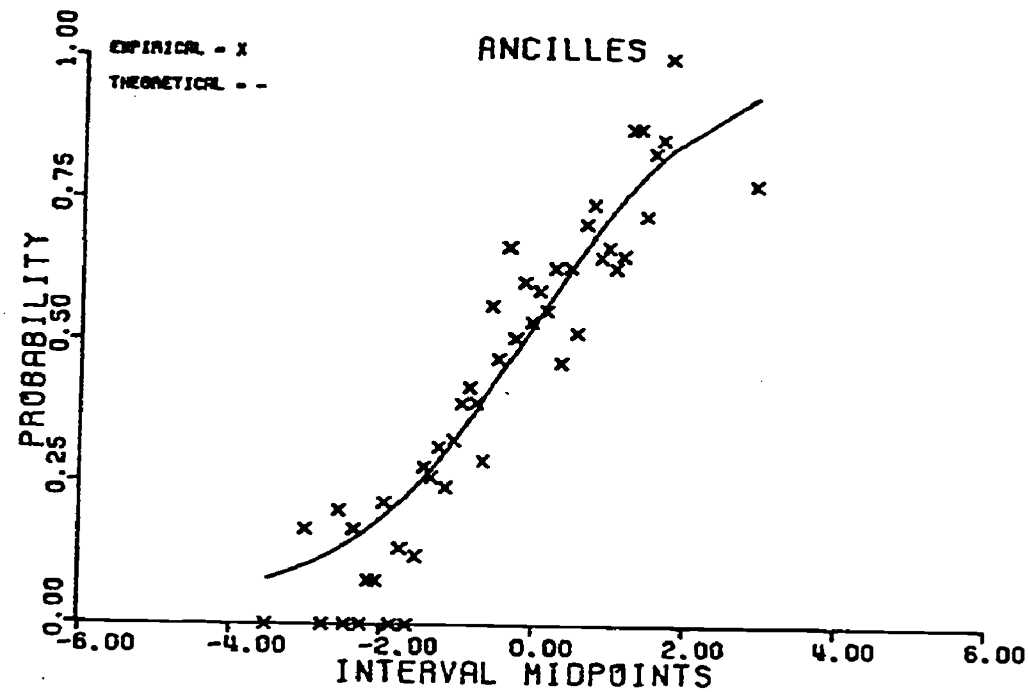
FIGURE 27

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 29



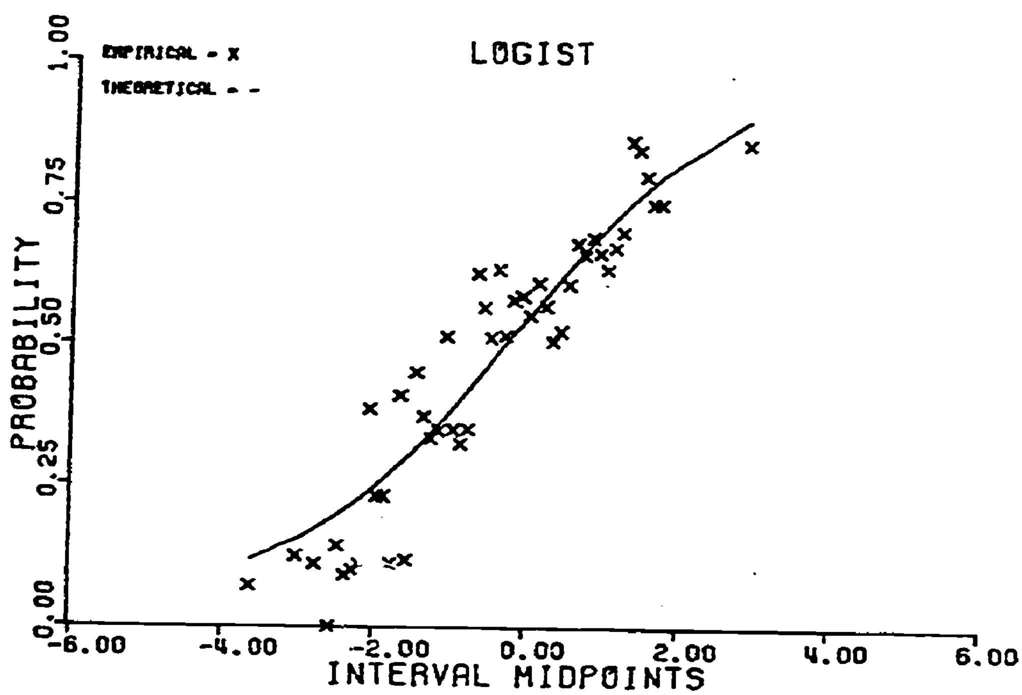


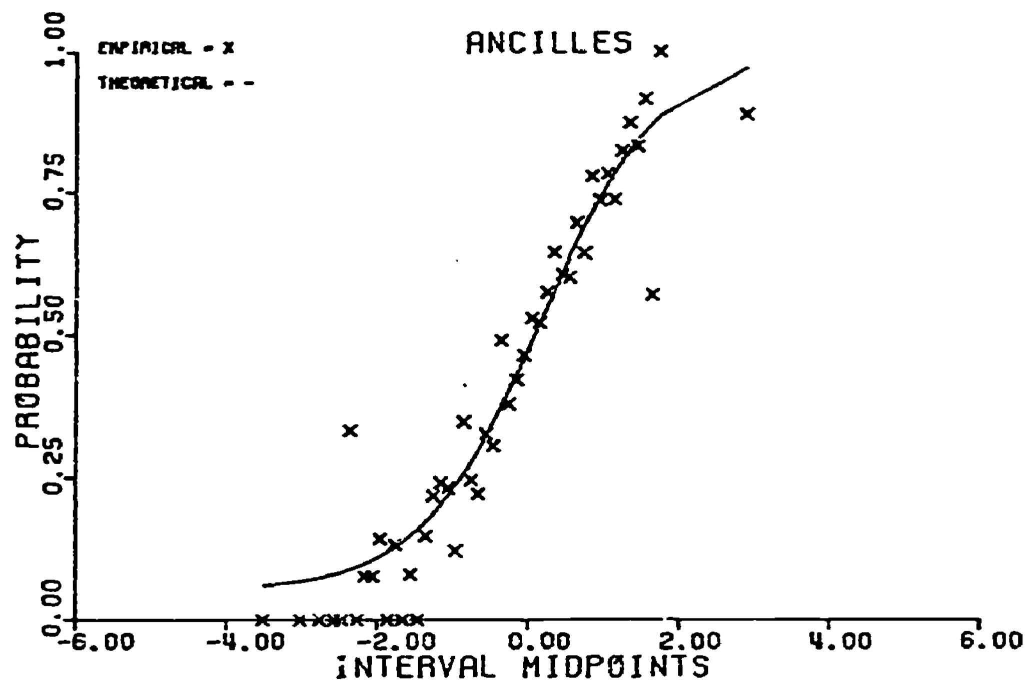
FIGURE 28

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 30



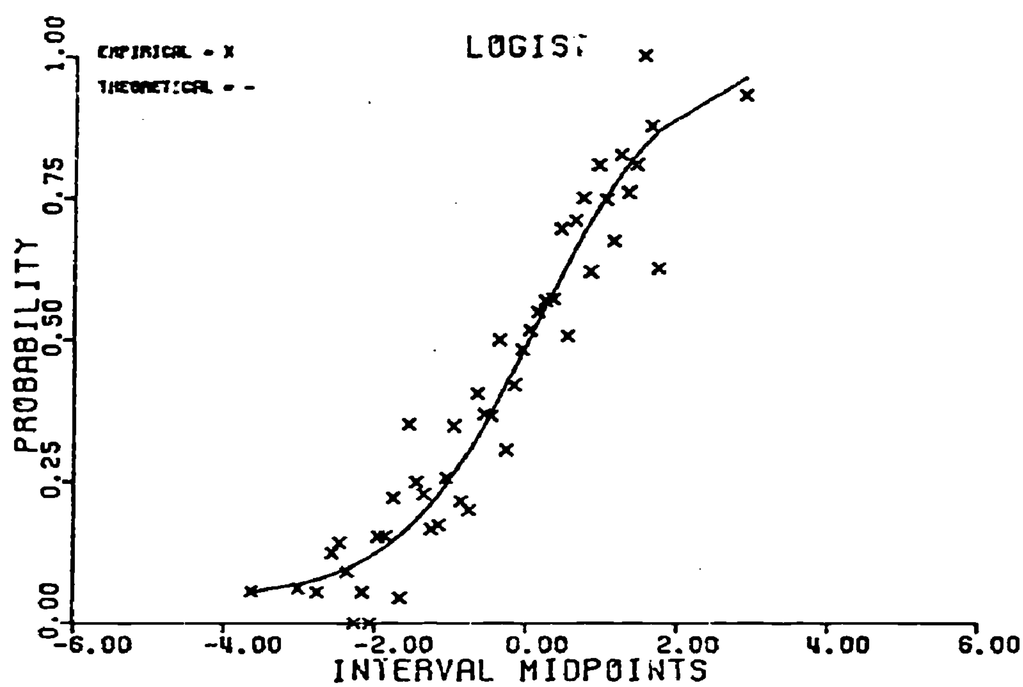
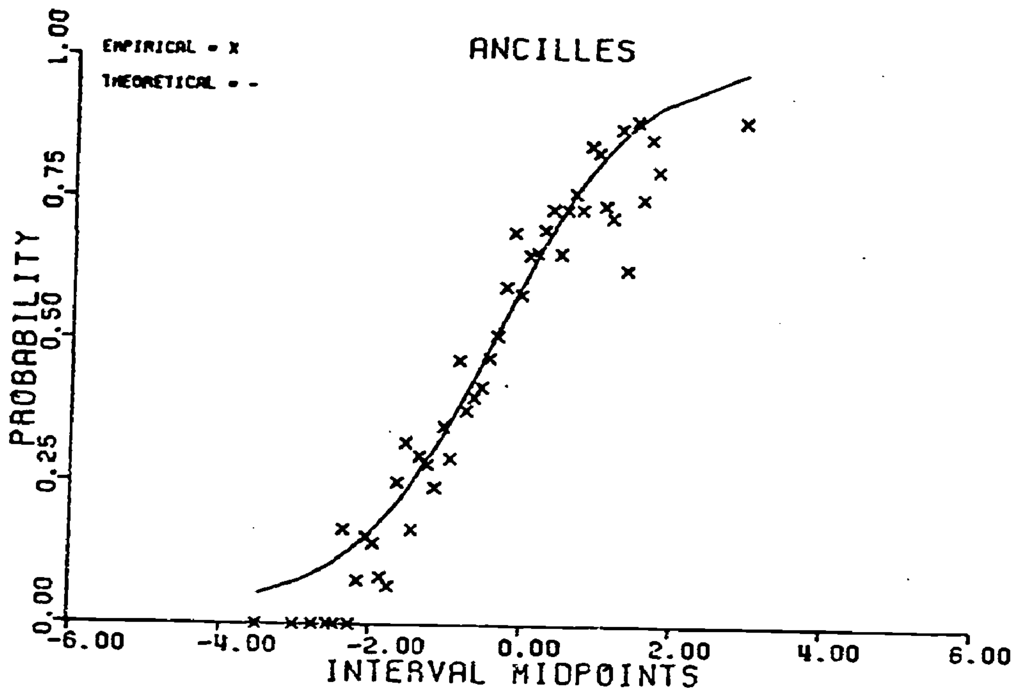


FIGURE 29

PLOTS OF EMPIRICAL AND
THEORETICAL CURVES BASED ON
ANCILLES AND LOGIST PROGRAMS

ITEM 31



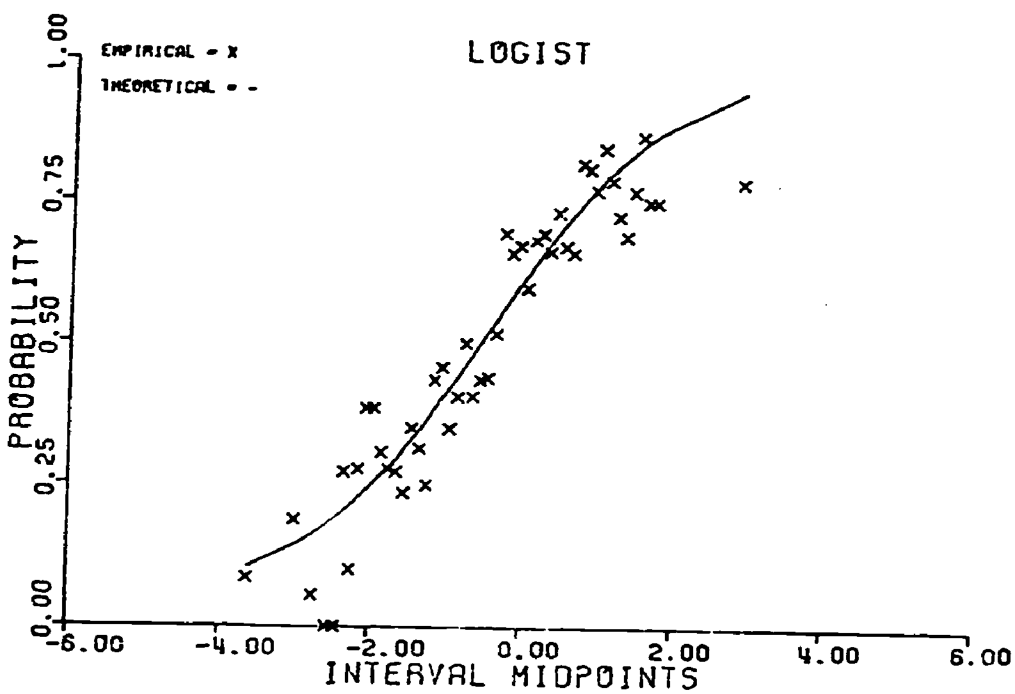


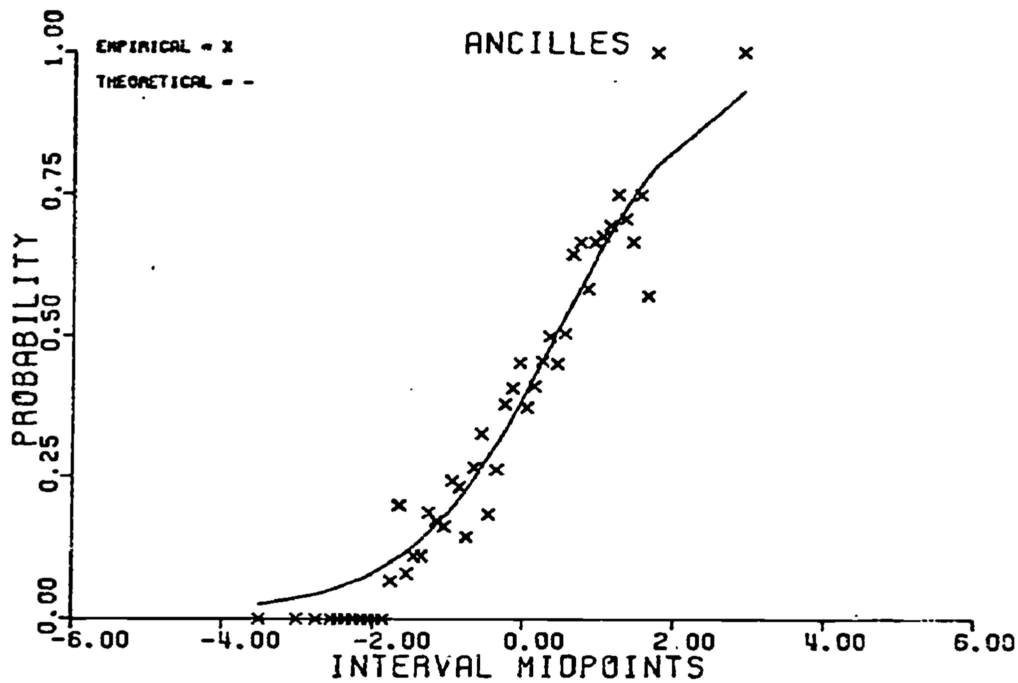
FIGURE 30

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 32



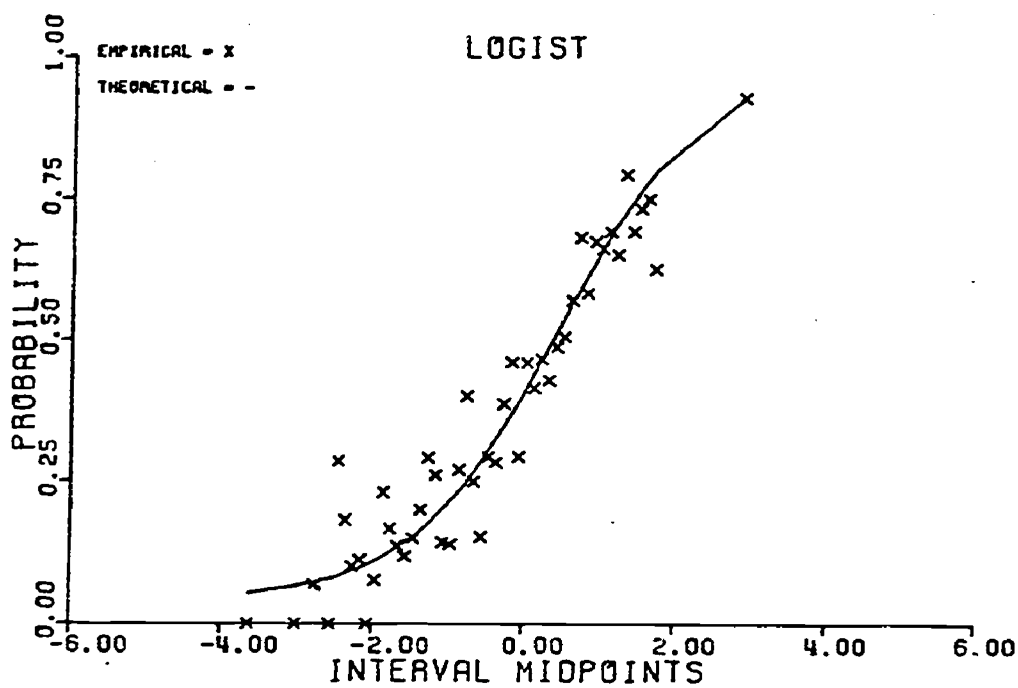


FIGURE 31

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 33

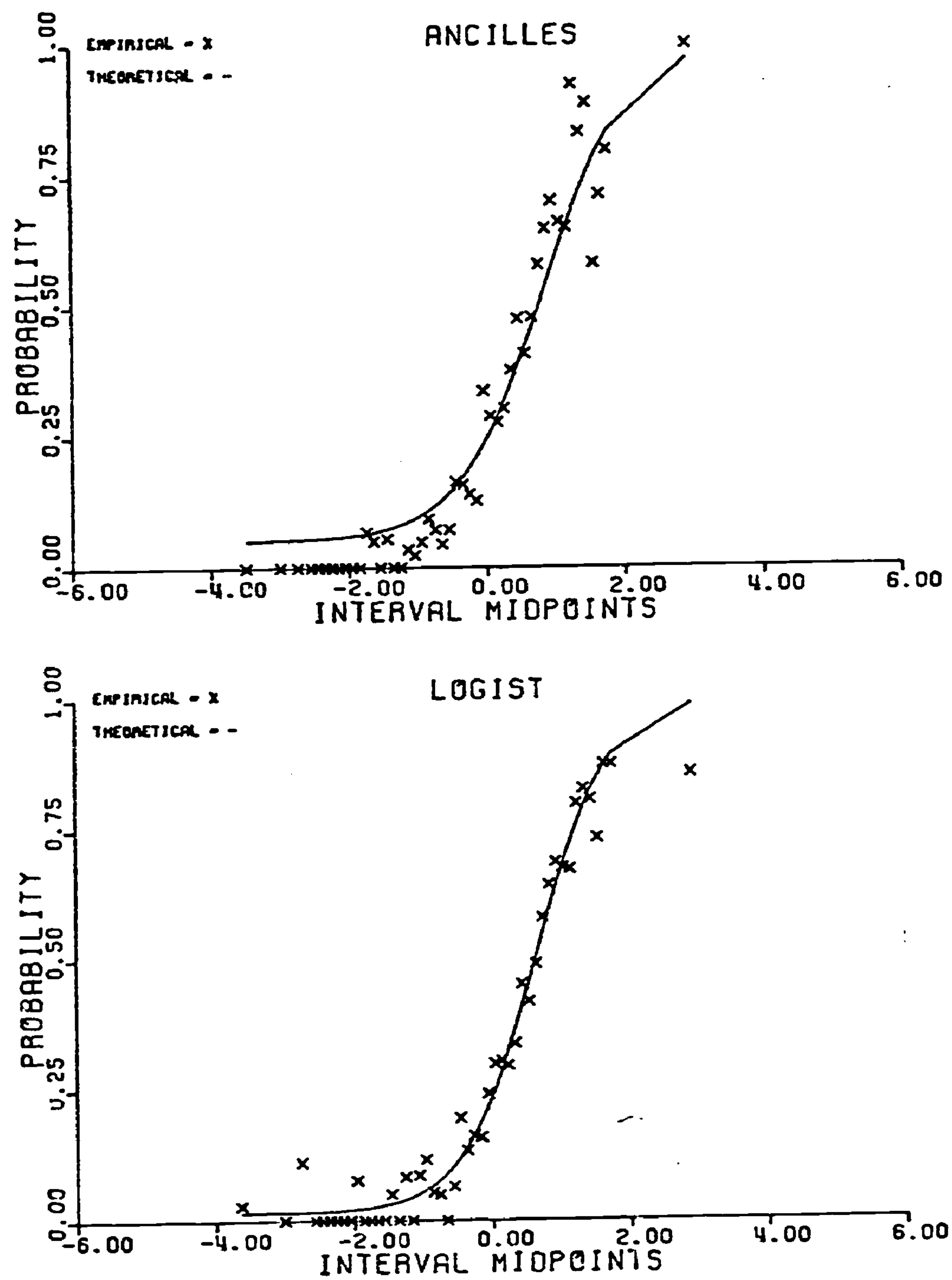


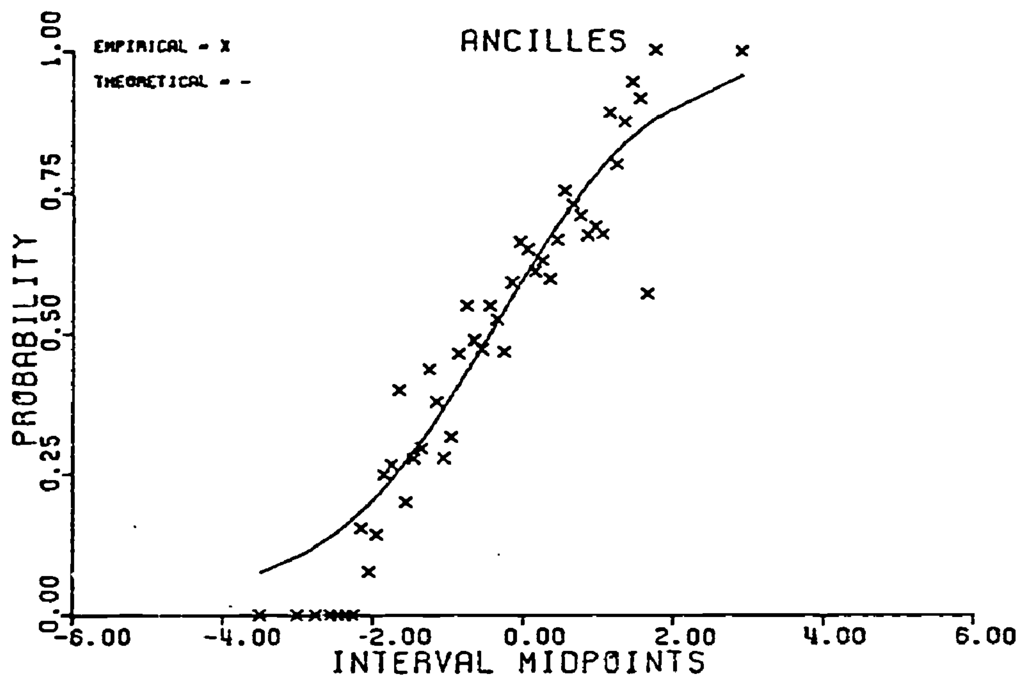
FIGURE 32

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 34



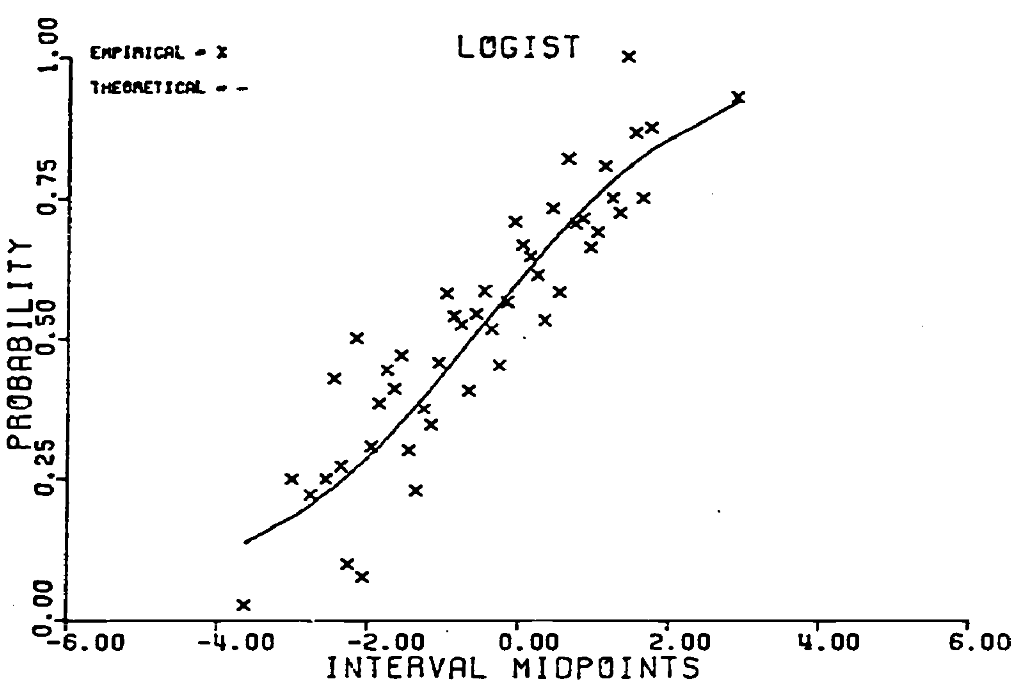


FIGURE 33

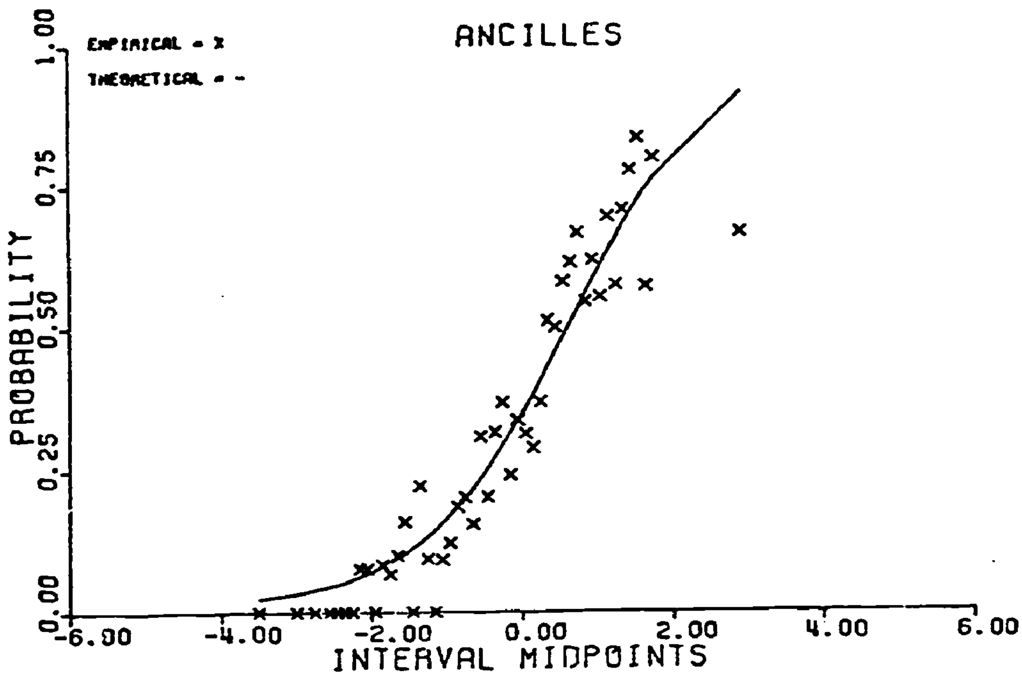
PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 35

ANCILLES



LOGIST

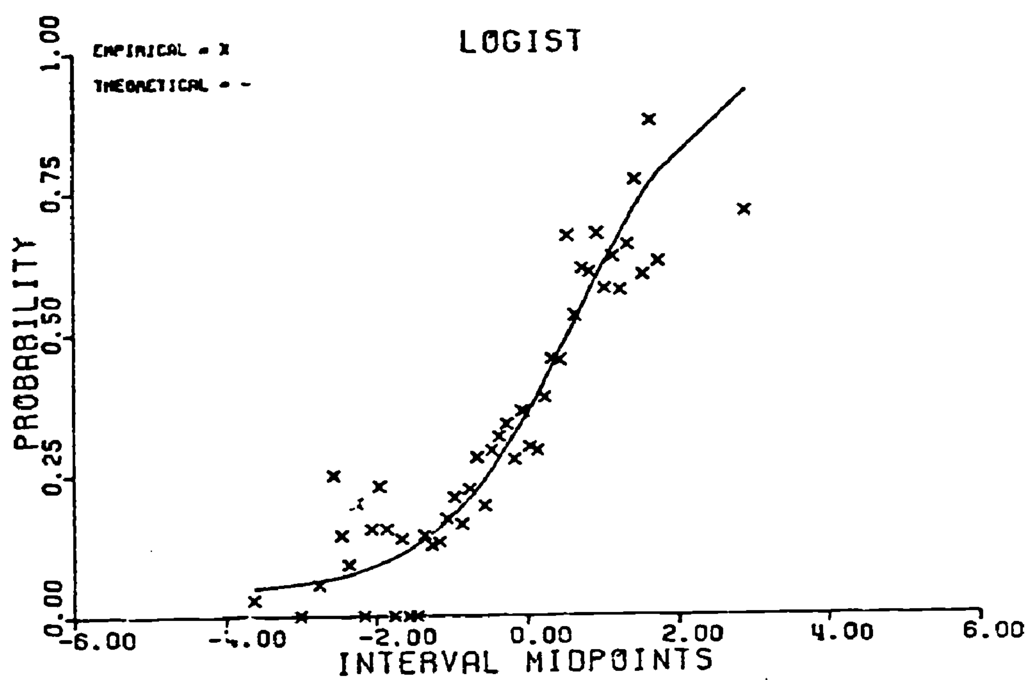


FIGURE 34

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 36

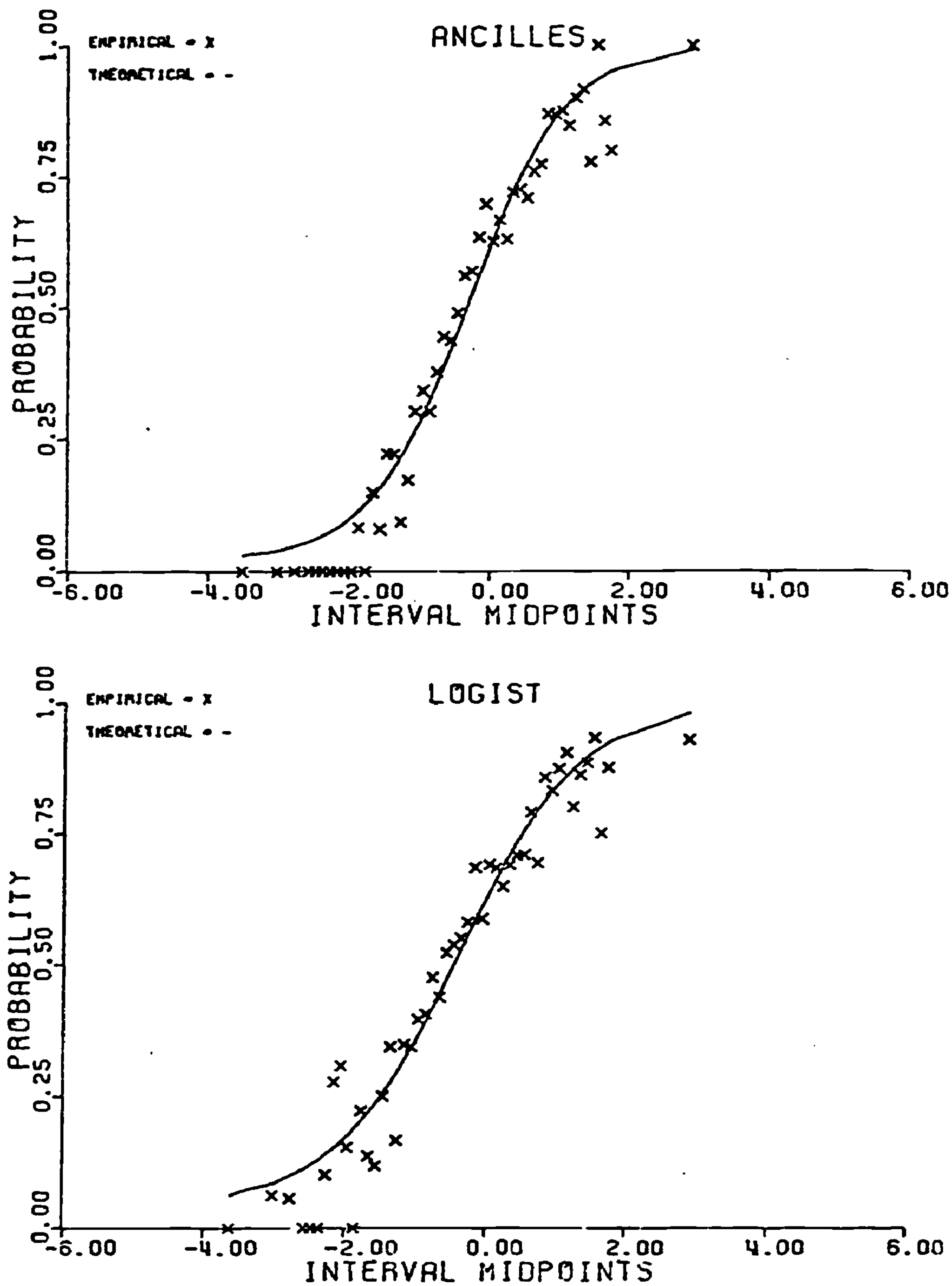


FIGURE 35

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 37

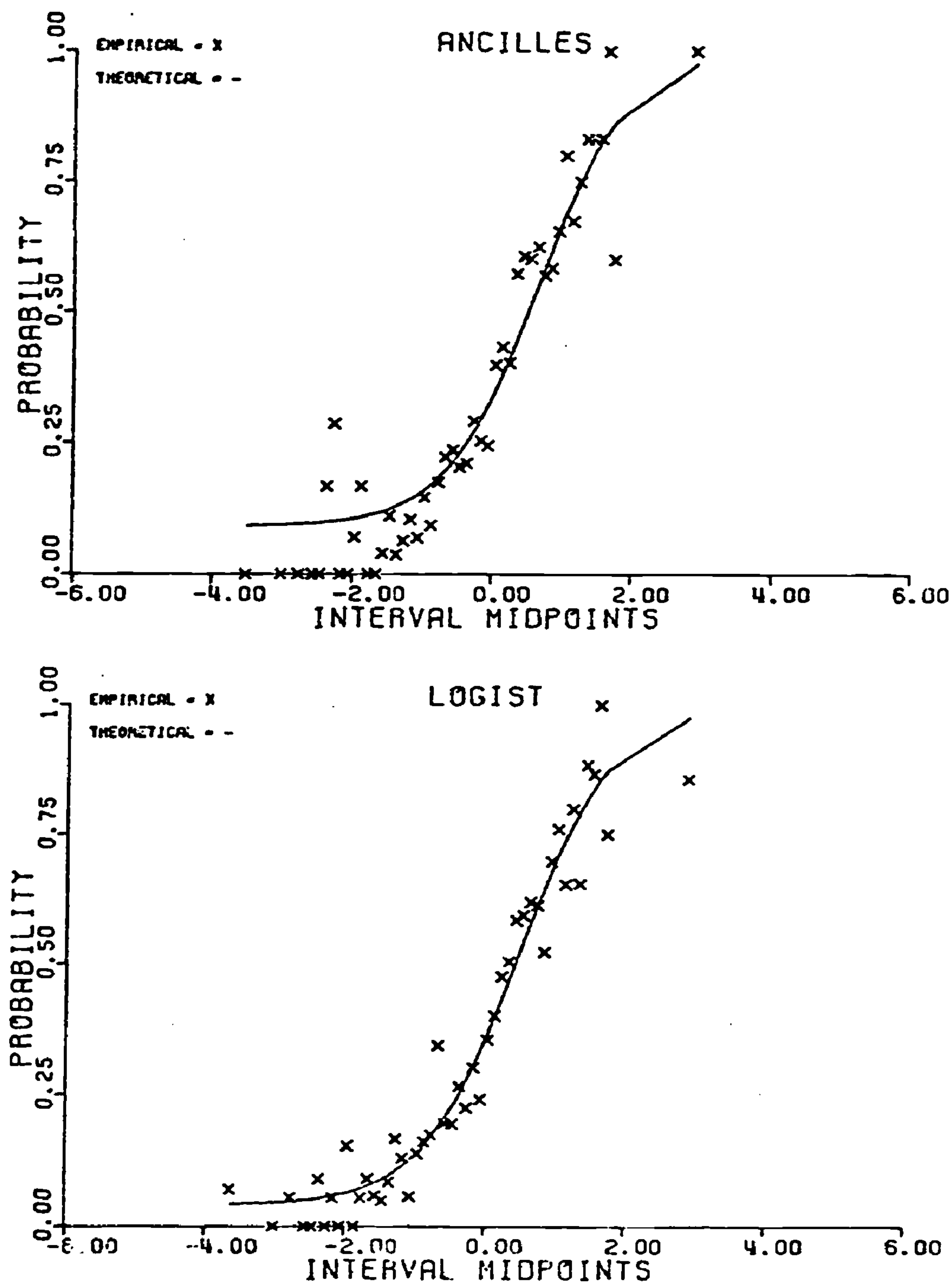


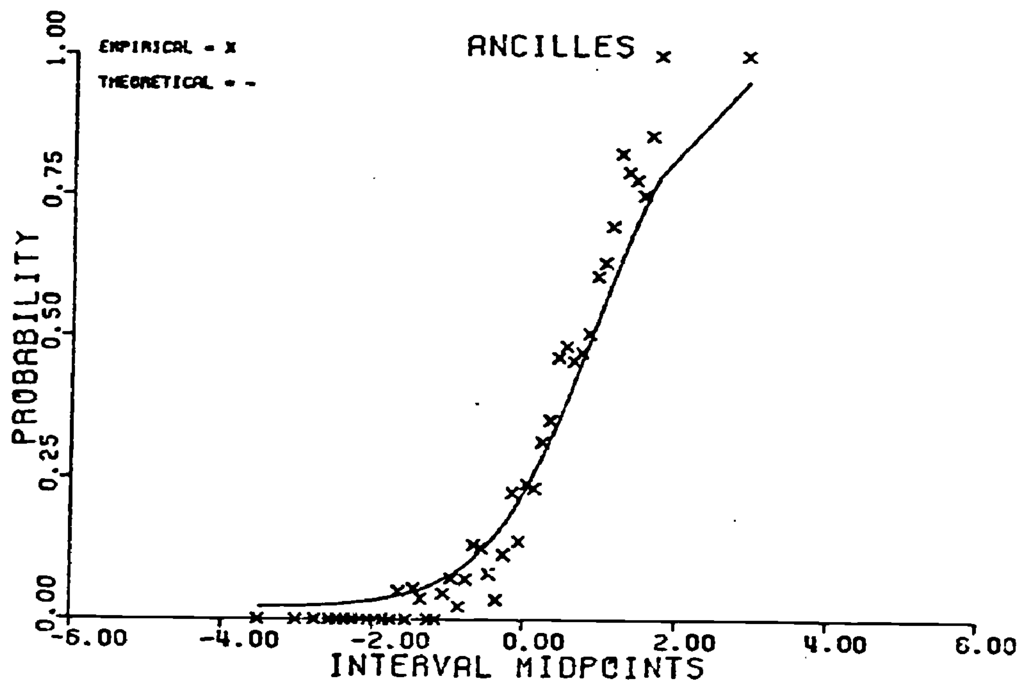
FIGURE 36

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 38



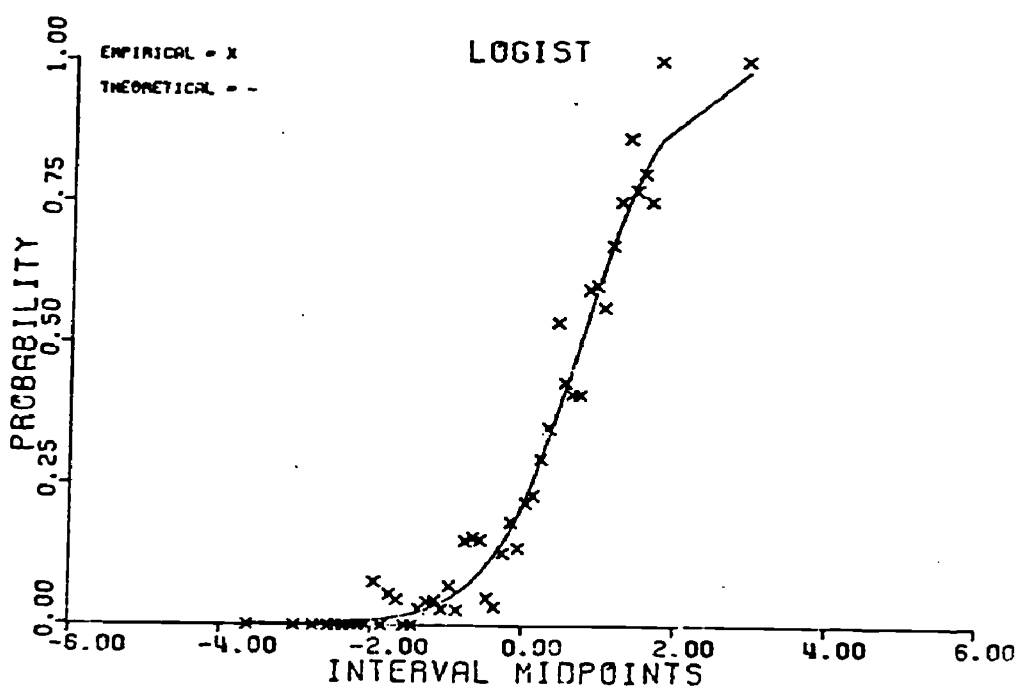


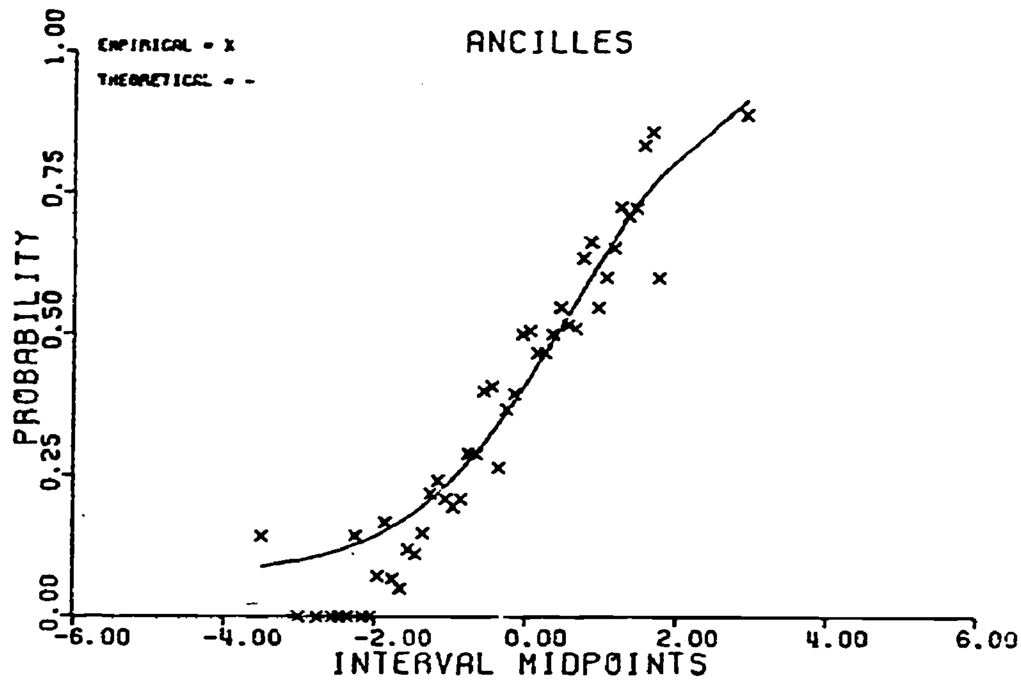
FIGURE 37

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 39



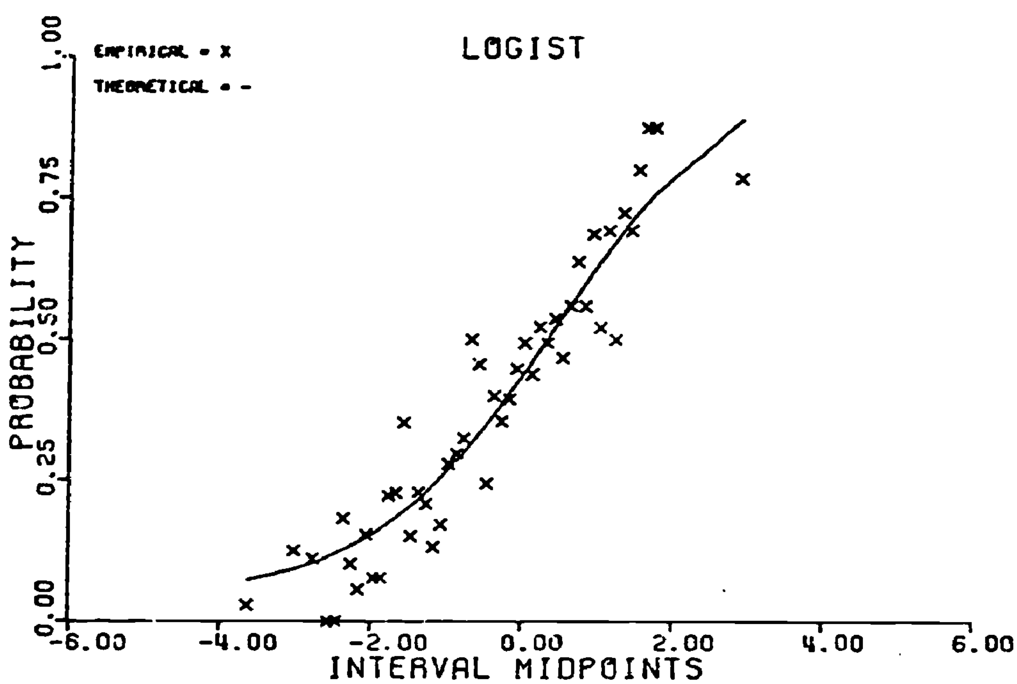


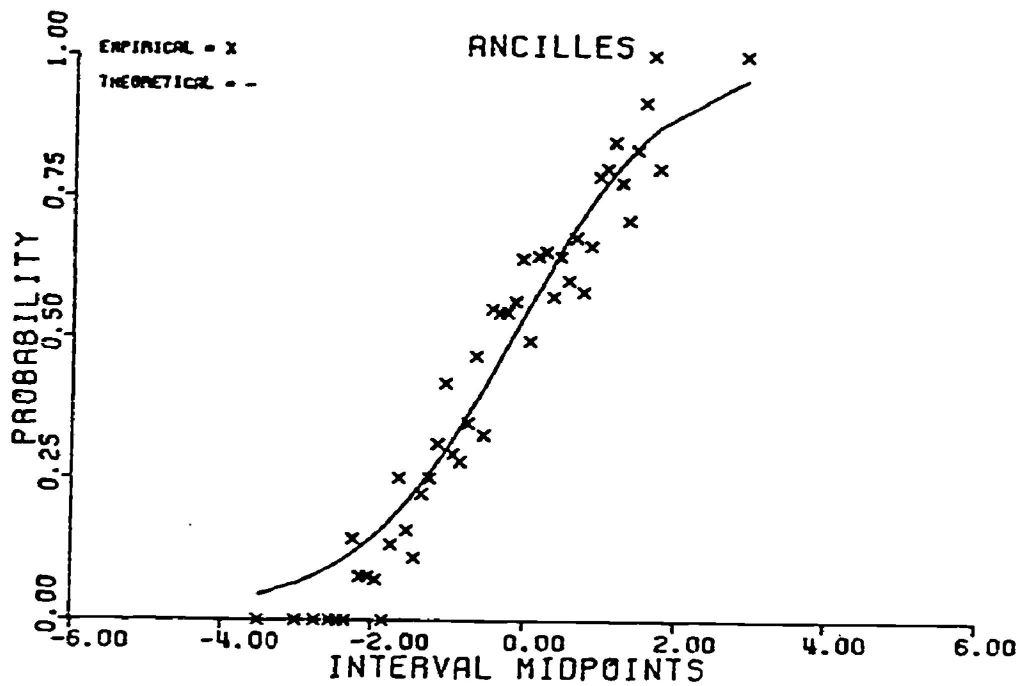
FIGURE 38

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 40



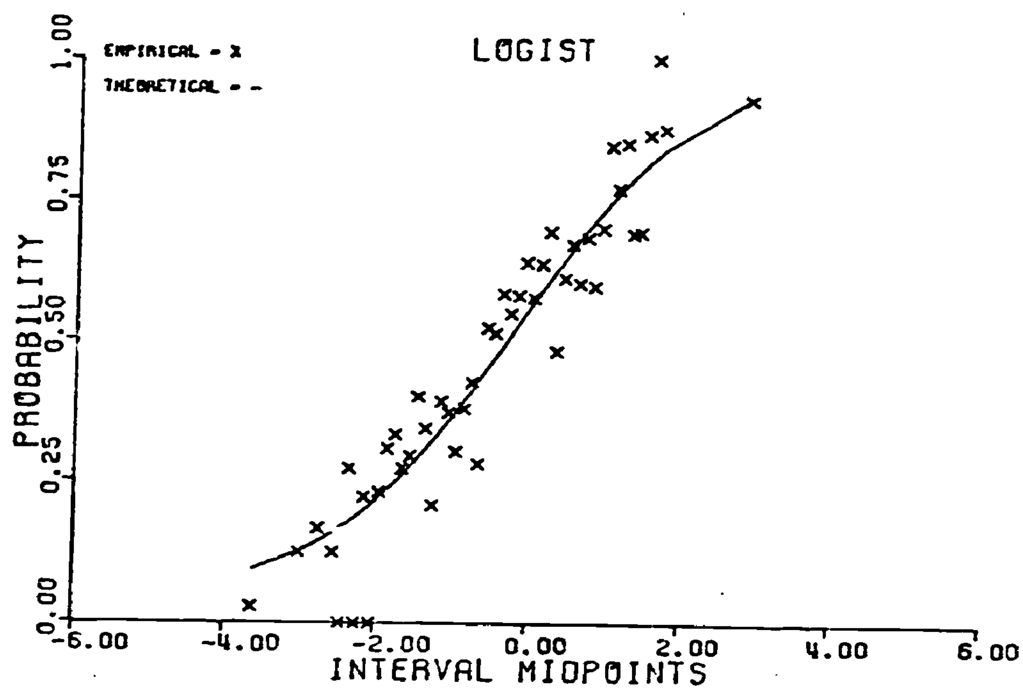


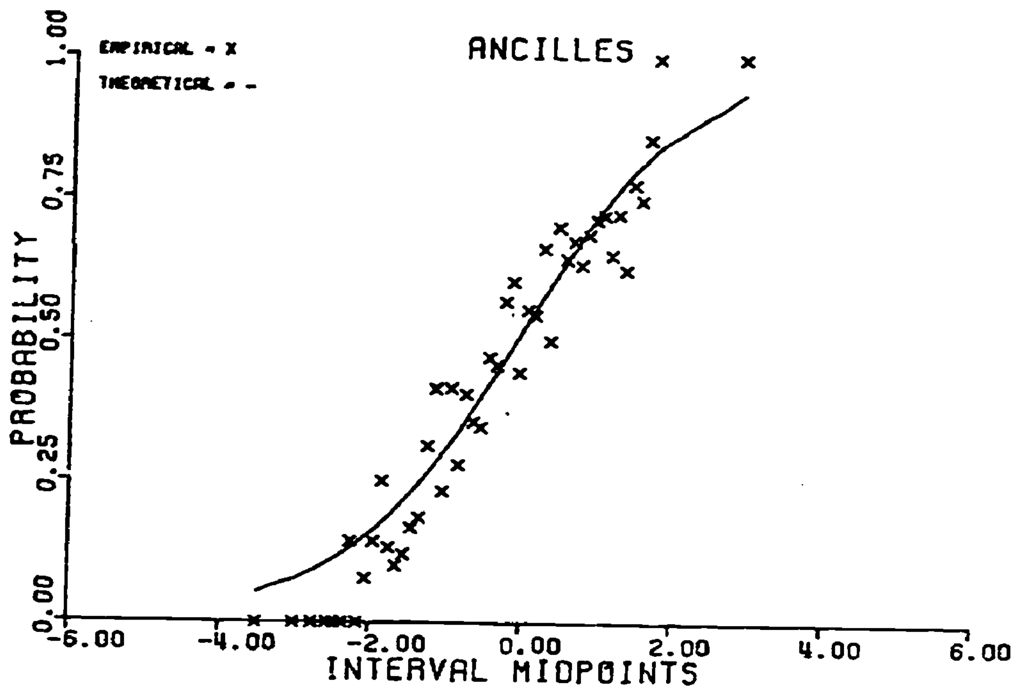
FIGURE 39

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 41



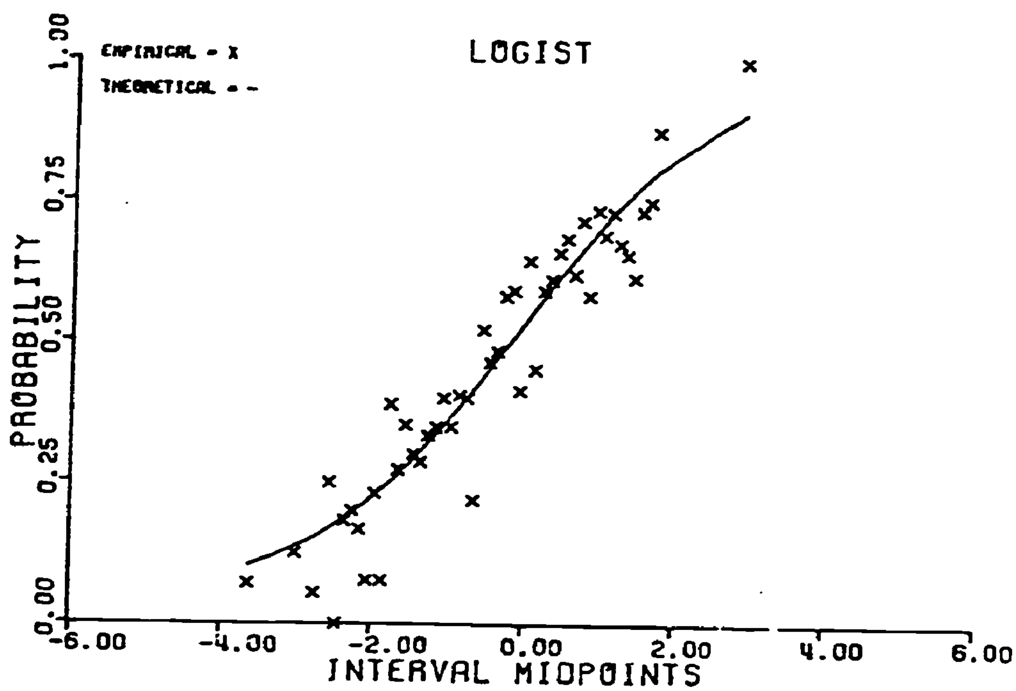


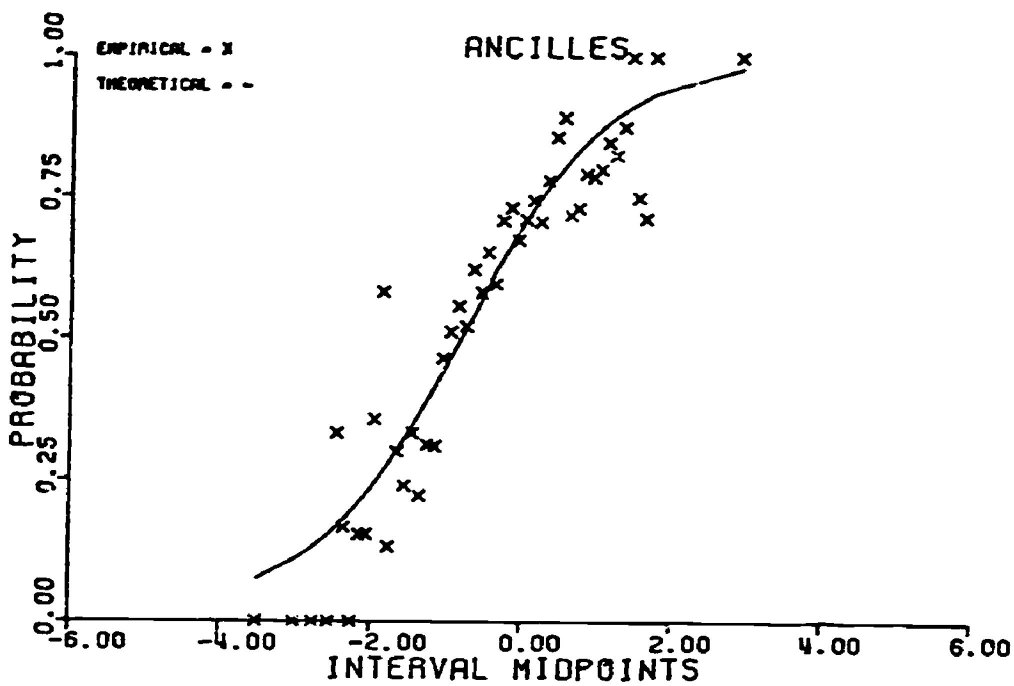
FIGURE 40

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 42



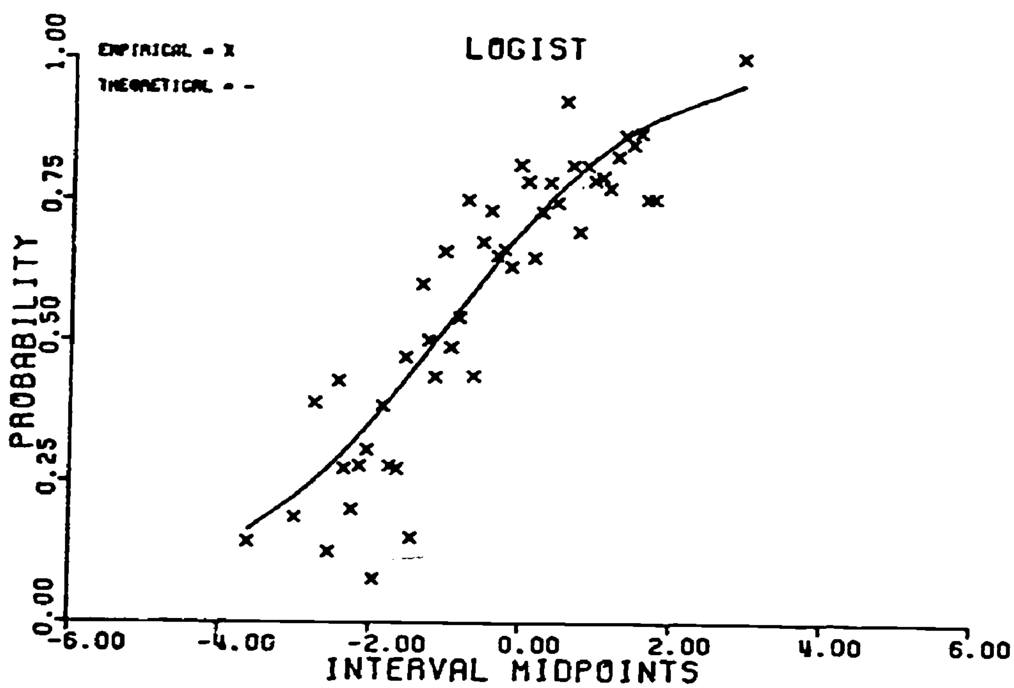


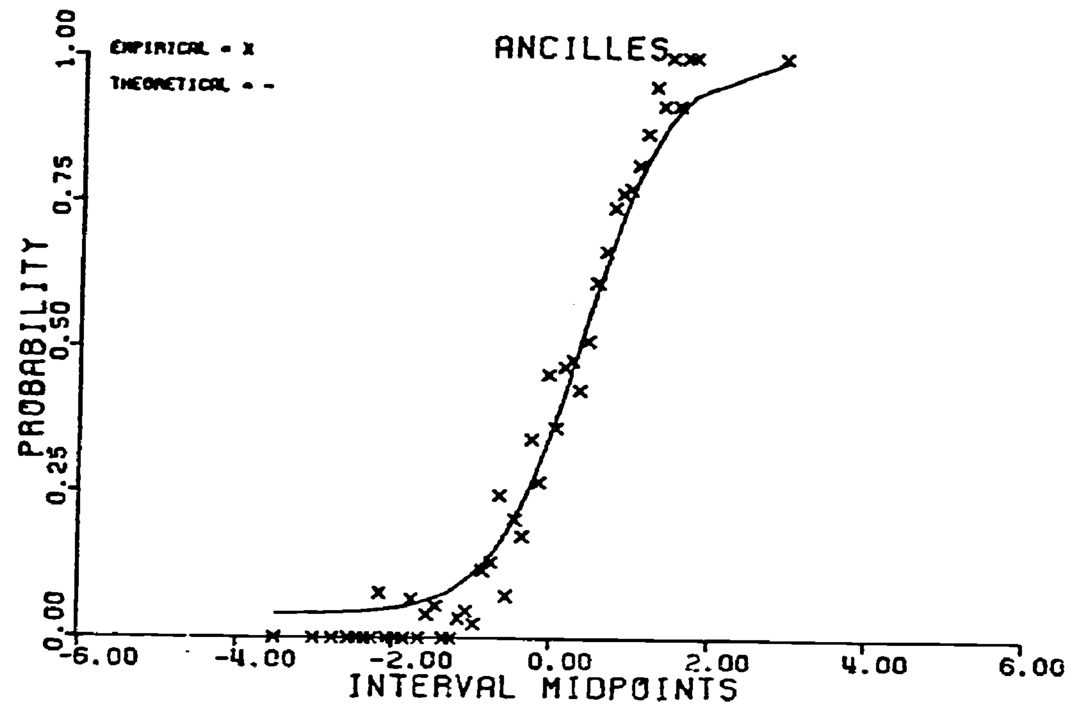
FIGURE 41

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 43



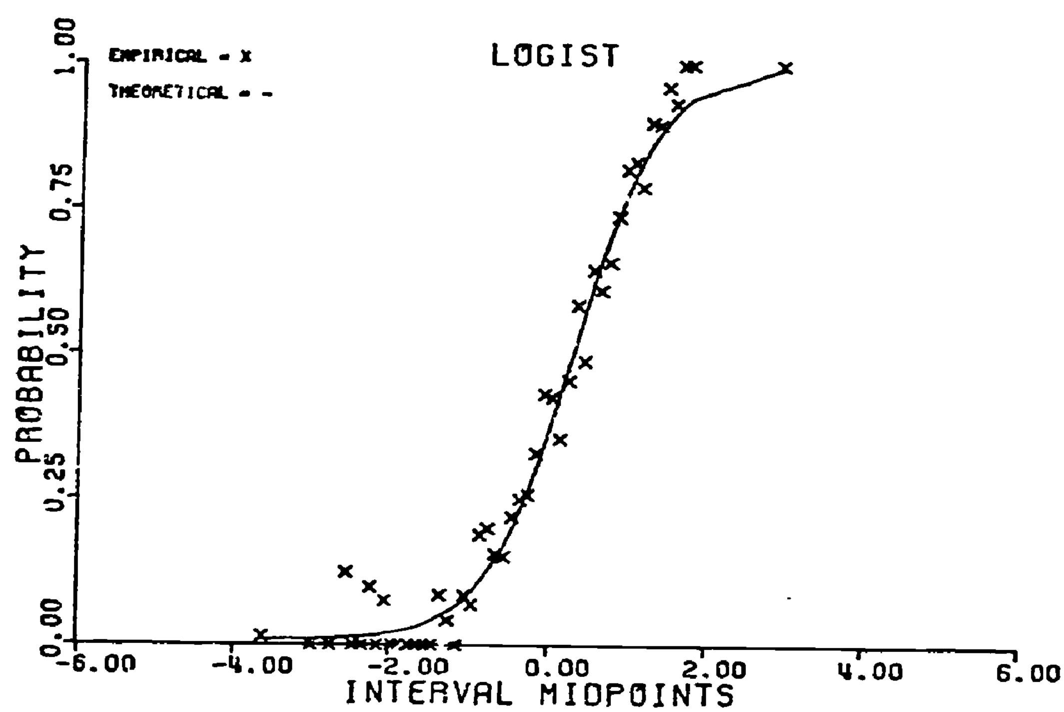


FIGURE 42

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 44

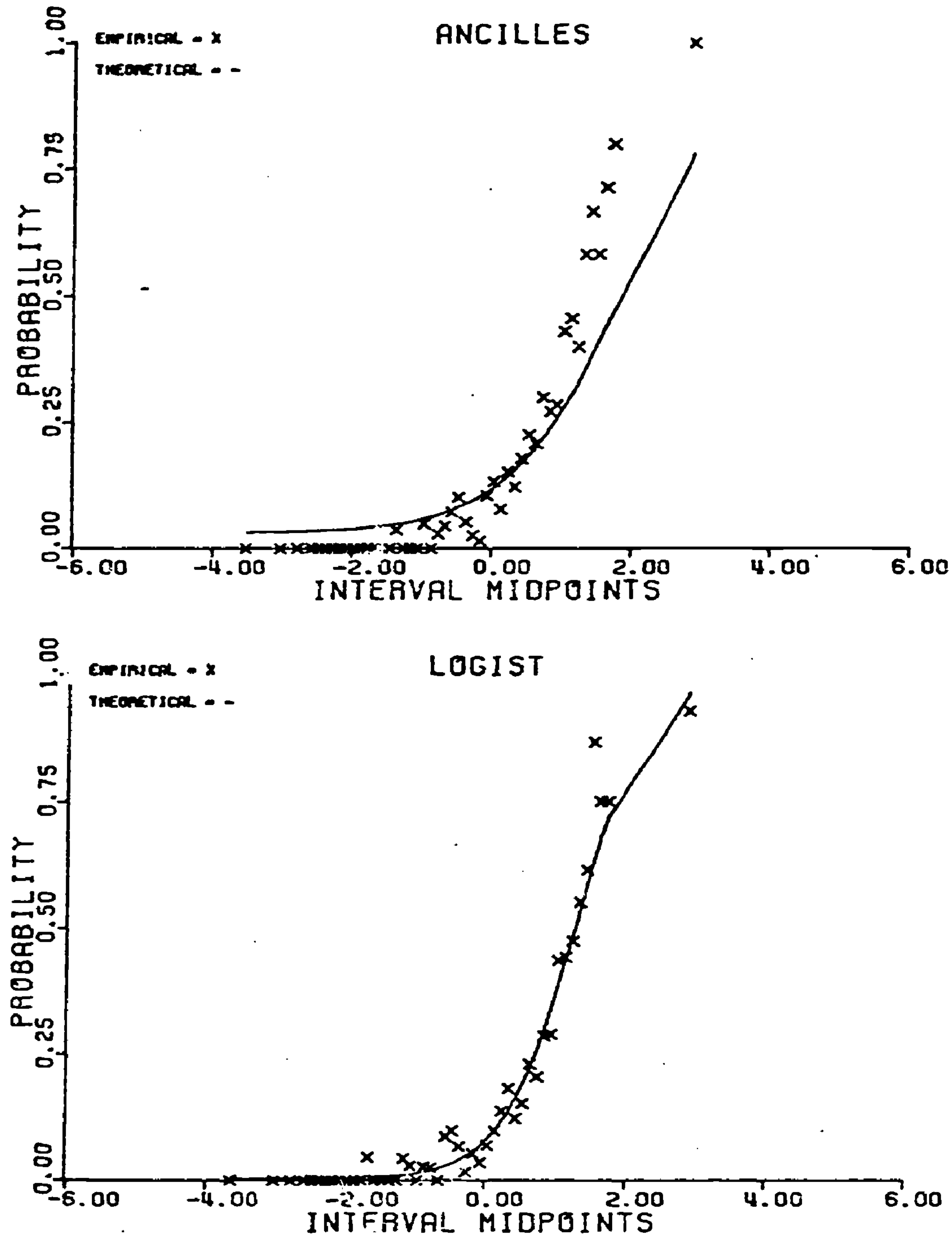


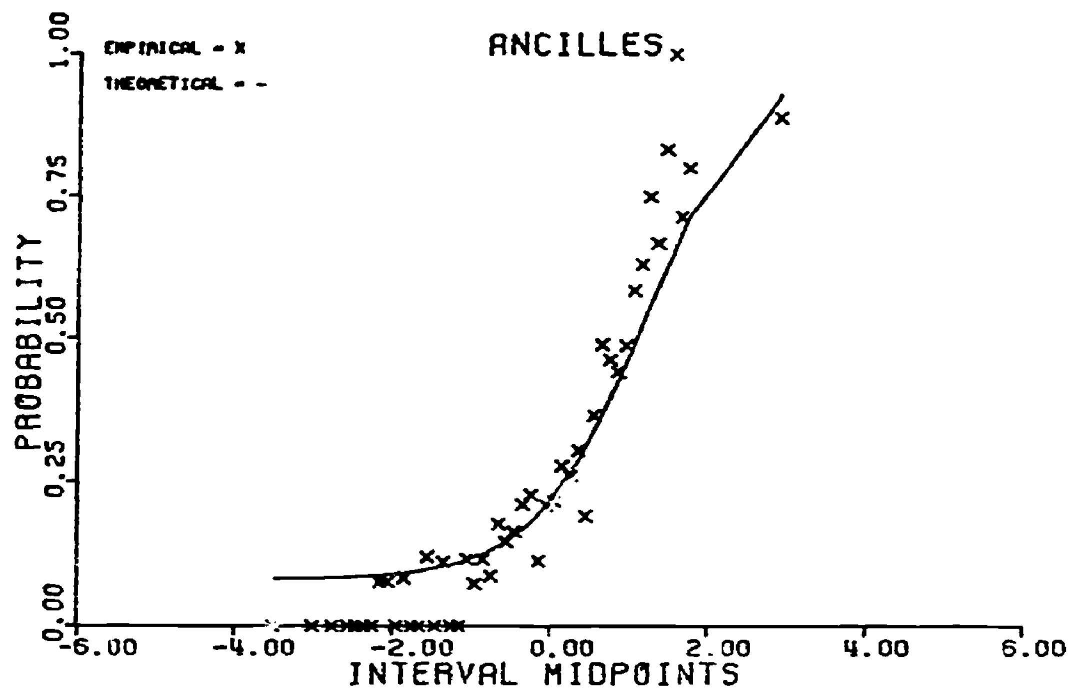
FIGURE 43

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 45



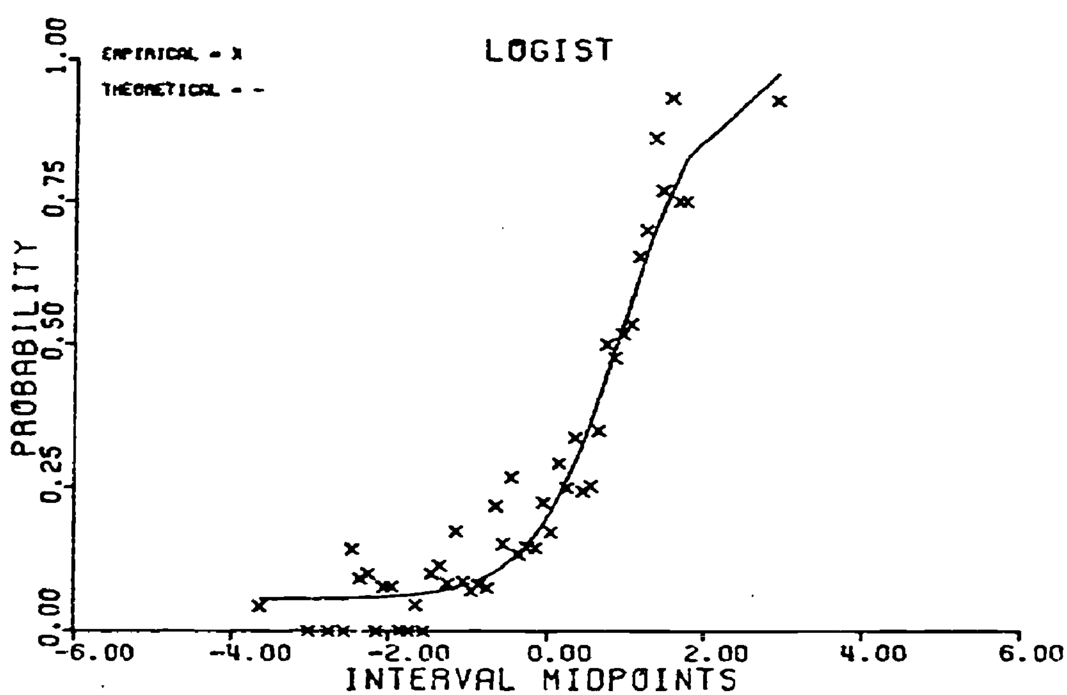


FIGURE 44

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 46

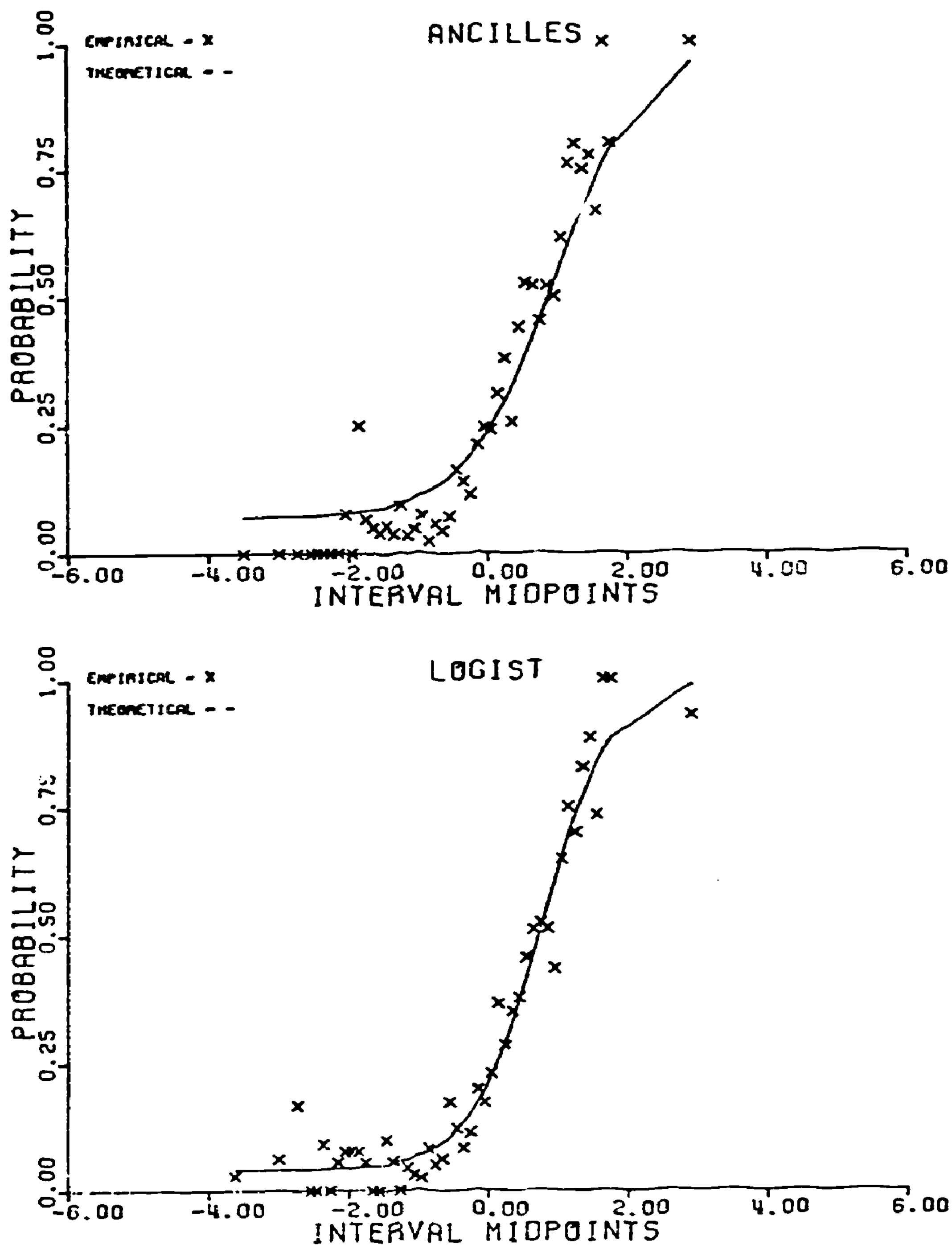


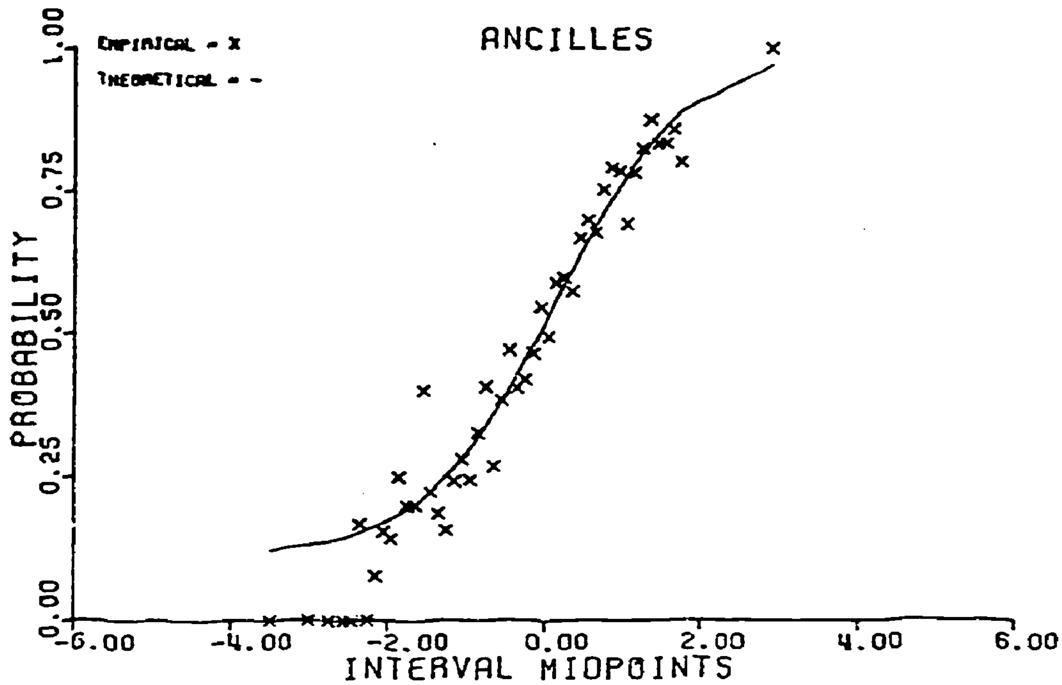
FIGURE 45

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 47



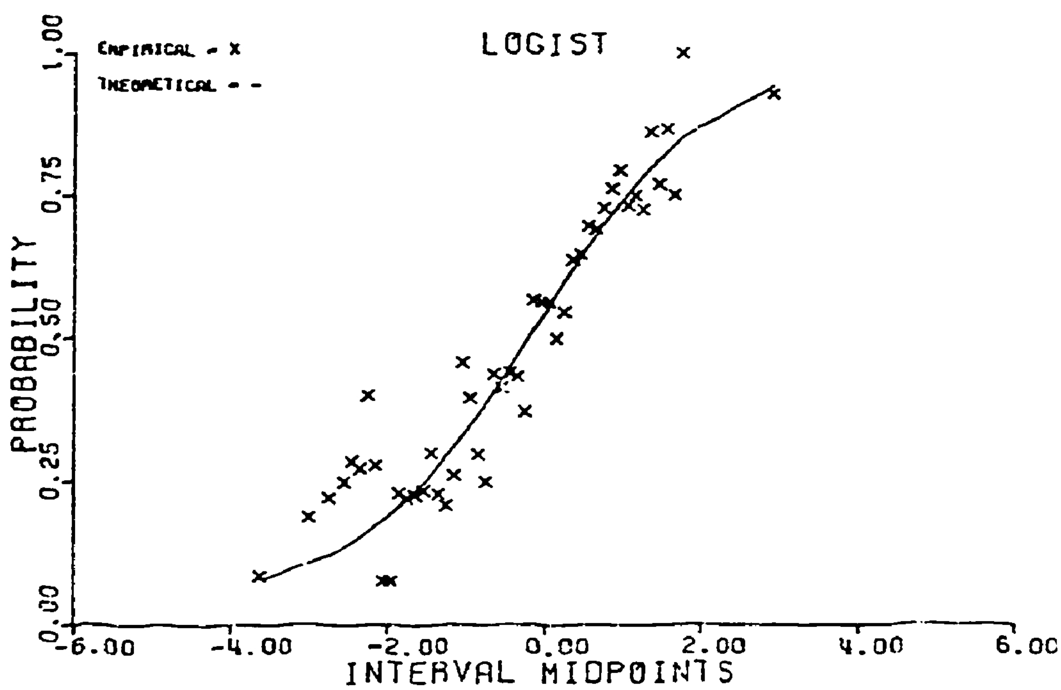


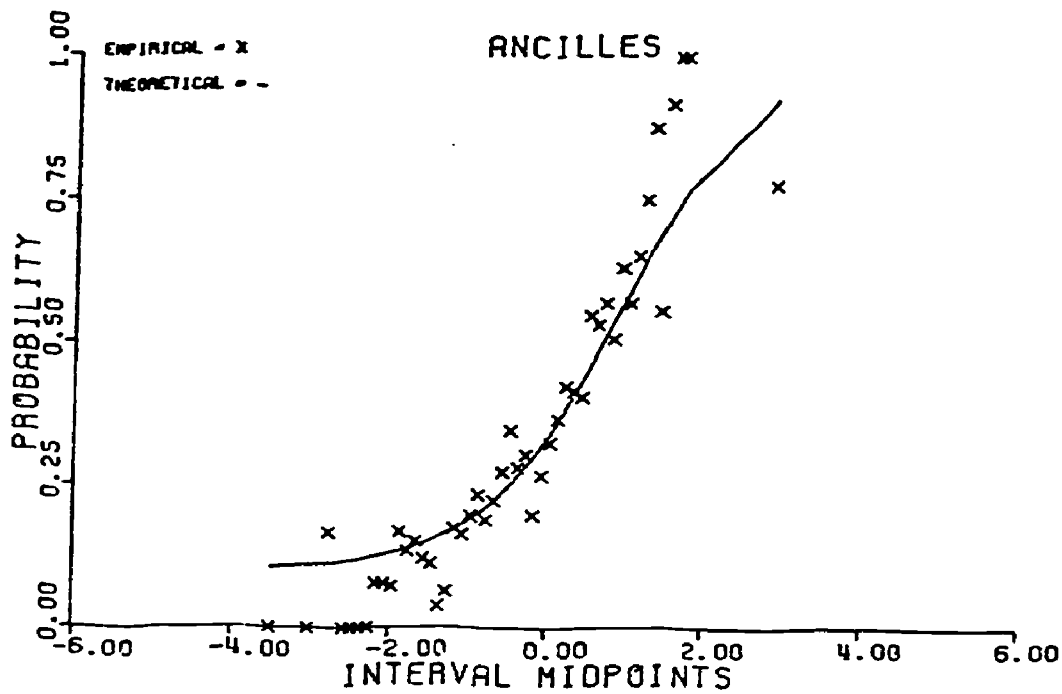
FIGURE 46

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 48



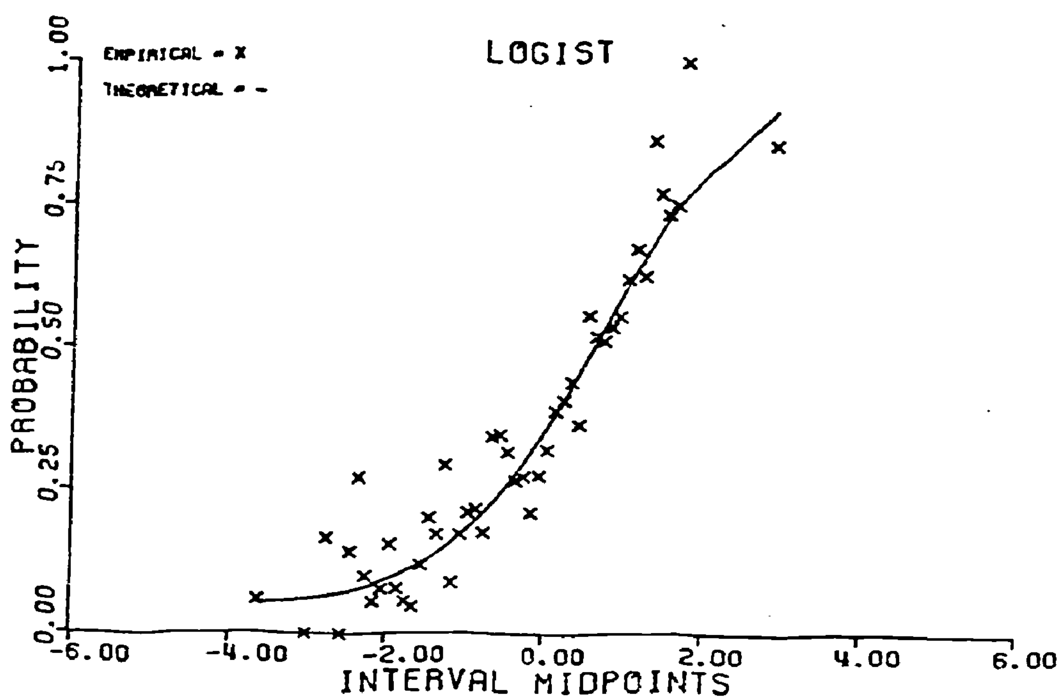


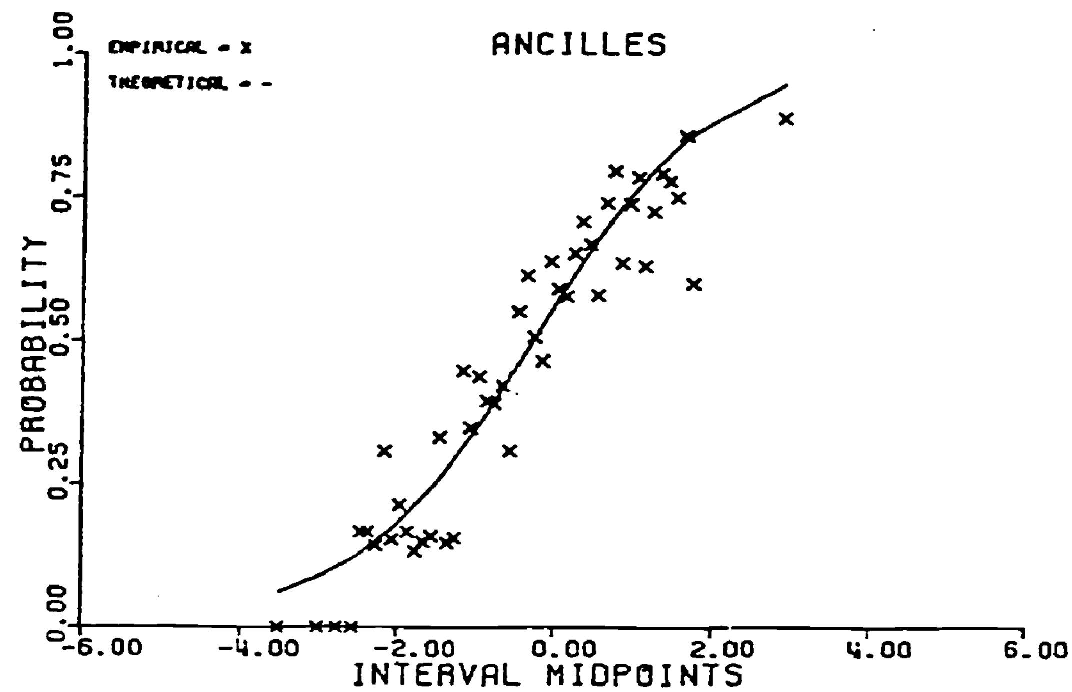
FIGURE 47

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 49



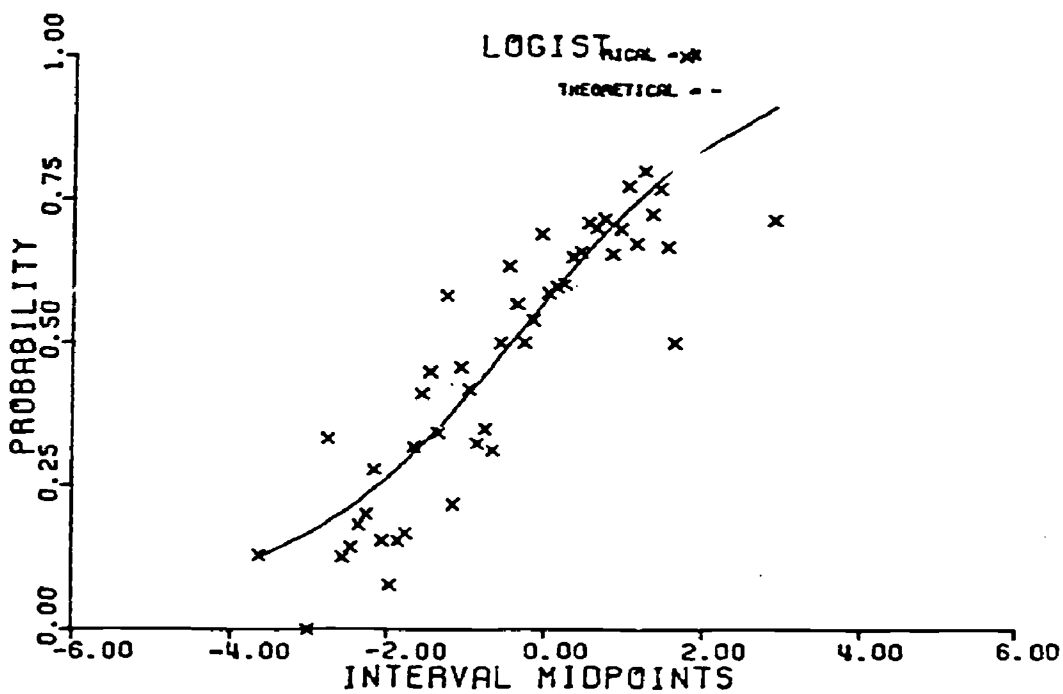


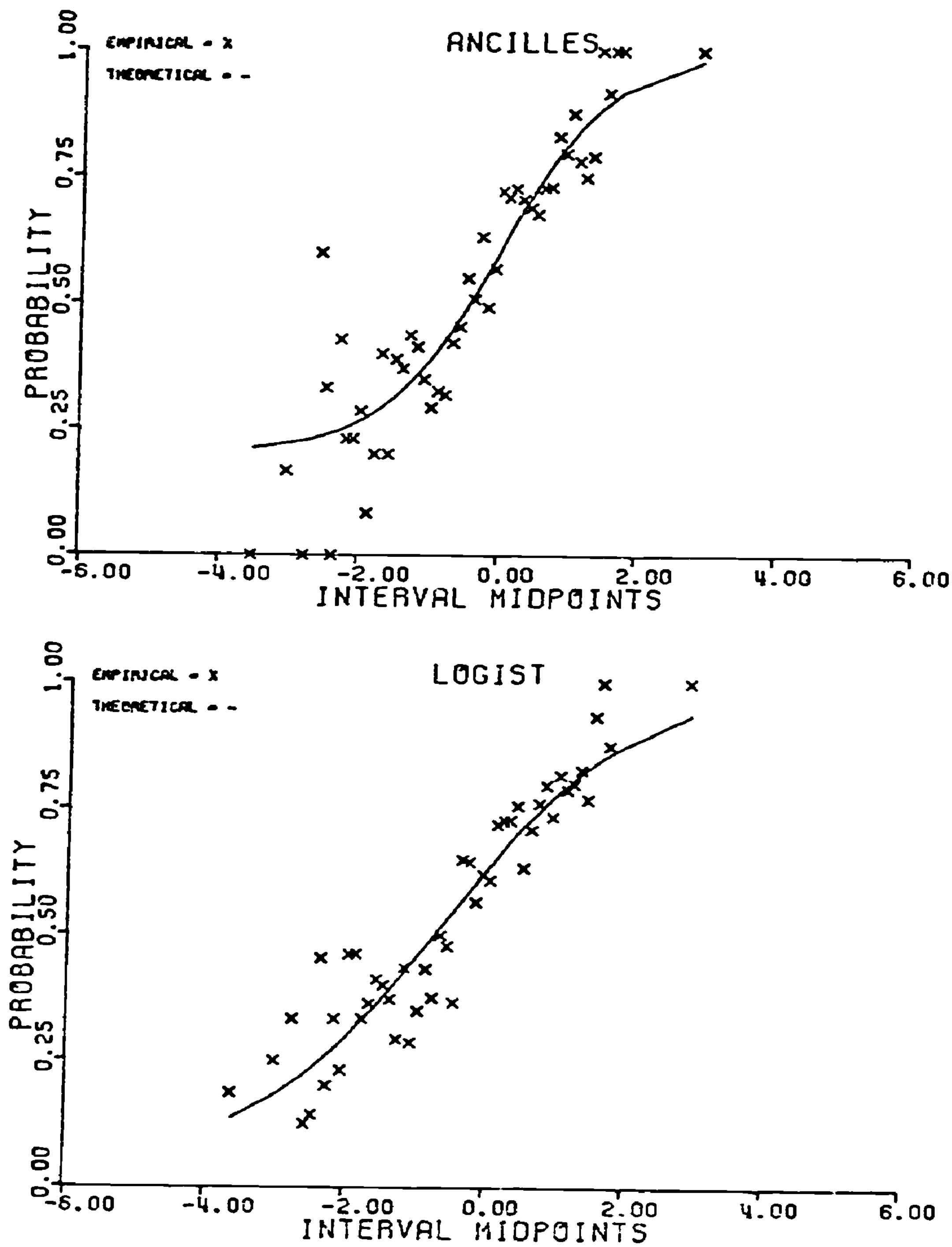
FIGURE 48

PLOTS OF EMPIRICAL AND

THEORETICAL CURVES BASED ON

ANCILLES AND LOGIST PROGRAMS

ITEM 50



Another analysis performed on the obtained chi-square values was to plot the distributions of chi-squares obtained for the ANCILLES and LOGIST procedures against the theoretical chi-square distribution for 45 degrees of freedom. These plots are shown in Figure 49 and Figure 50 for the ANCILLES and LOGIST procedures, respectively. From these plots it is clear that the chi-squares obtained for the ANCILLES procedure were shifted to the right from the expected distribution. The LOGIST chi-square distribution was also shifted somewhat to the right, but not nearly so much as the ANCILLES chi-squares.

One final analysis performed on the chi-square values was to perform a chi-square test of independence for the two procedures. That is, using the obtained chi-square values, items were classified as fitting or nonfitting for each of the two procedures. A chi-square test was then performed to test whether the classification using chi-squares for ANCILLES was independent of classification using the LOGIST chi-squares. A chi-square value of 3.43 was obtained. The critical value for $\alpha = .05$ was $\chi^2(1) = 3.84$, so the hypothesis of independence was not rejected. There was apparently no association in the items categorized as fitting or nonfitting between the two methods of classification. This result was supported by the results of a test for the significance of a coefficient of agreement. A kappa coefficient (Cohen, 1960) was computed on the chi-square classifications, and the kappa was then converted to a $\underline{z}$-score. A kappa equal to .228 was obtained, and a $\underline{z} = 1.2$ resulted from dividing the kappa coefficient by its standard error of measurement ($\sigma_k = .19$). The null hypothesis of no agreement was not rejected.

MSD Statistics

The MSD statistics obtained for the two procedures are displayed in Table 2. The dependent $\underline{t}$-test performed on these values showed the mean ANCILLES MSD value to be significantly higher than the mean LOGIST MSD value ($\underline{p} < .05$). However, a comparison of Table 2 with Table 1 indicates that there is no apparent relationship between the size of the chi-square values and the MSD statistics obtained for the items for either procedure. A Pearson product moment correlation was computed for the MSD and chi-square values and the correlations for both the LOGIST and ANCILLES procedures were found to be not significantly different from zero ($\underline{r} = .12$ for ANCILLES and $\underline{r} = .19$ for LOGIST).

FIGURE 49
OBSERVED DISTRIBUTION OF
CHI SQUARES FOR ANCILLES
WITH EXPECTED CHI
SQUARE DISTRIBUTION

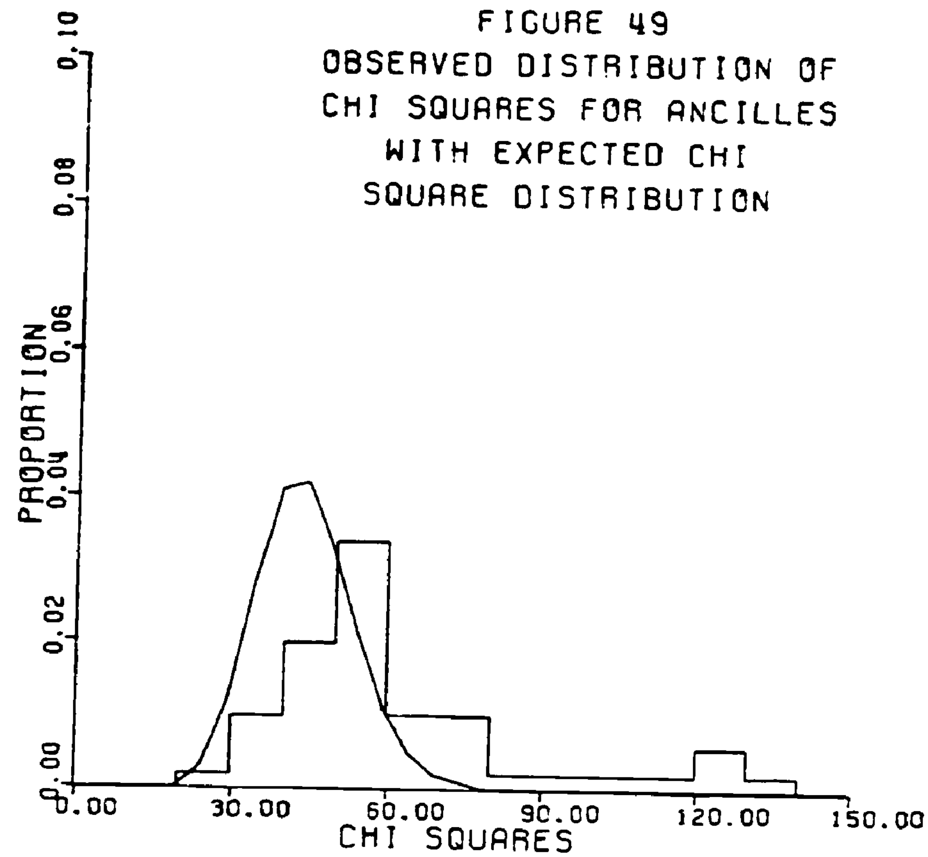


FIGURE 50
OBSERVED DISTRIBUTION OF
CHI SQUARES FOR LOGIST
WITH EXPECTED CHI
SQUARE DISTRIBUTION

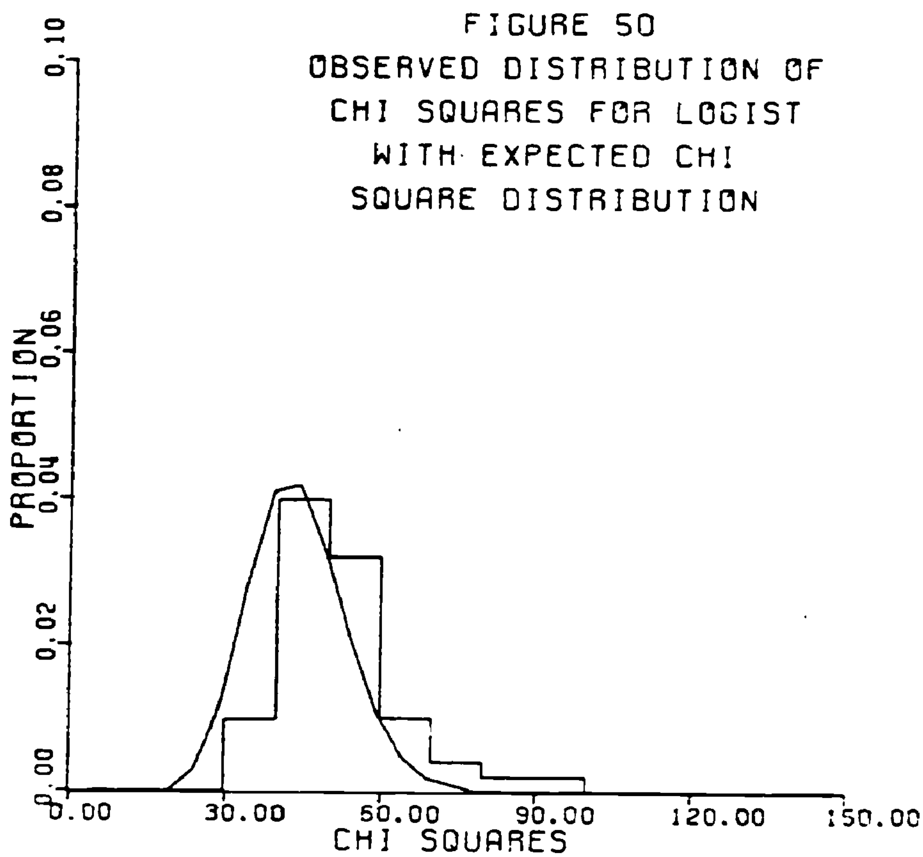


Table 2
ANCILLES vs. LOGIST
Goodness of Fit Comparison
Using the MSD Statistic

Item	ANCILLES MSD	LOGIST MSD
1	---	.233
2	.202	.198
3	.179	.173
4	.186	.186
5	.158	.158
6	.160	.159
7	.178	.176
8	.212	.213
9	---	.191
10	.209	.208
11	.225	.226
12	.181	.179
13	.195	.195
14	.215	.215
15	.156	.159
16	.191	.192
17	.209	.210
18	.220	.220
19	.185	.184
20	.194	.194
21	.222	.223
22	.201	.203
23	.192	.190
24	.228	.229
25	.206	.205
26	.191	.192
27	.207	.208
28	.209	.209
29	.220	.220
30	.199	.199
31	.201	.203
32	.202	.202
33	.161	.154
34	.213	.213
35	.197	.196
36	.181	.182
37	.185	.182
38	.155	.150
39	.215	.214
40	.211	.212
41	.217	.218
42	.190	.191
43	.161	.159
44	.113	.102
45	.167	.160
46	.166	.157
47	.207	.208
48	.200	.198
49	.216	.217
50	.204	.207
$\bar{X}$ $\underline{t}(47) = 2.15$	.194 $(\underline{p} < .05)$	.193

Note: The critical value of $\underline{t}(47) = 2.014$ for $\alpha = .05$.

Parameter Estimate Distribution Analyses

Item Parameter Estimates The item parameter estimates obtained from ANCILLES and LOGIST are shown in Table 3. The correlations of the two sets of estimates are displayed in Table 4. Because the origin and unit of measurement used for the ability and item parameter estimates are arbitrary, the scales used for the two sets of estimates are different. Therefore, to facilitate this comparison the ANCILLES estimates were put on the same scale as the LOGIST estimates using procedures set out by Marco (1977). The scaled ANCILLES *a*- and *b*-values are presented in Table 5. Scaling does not alter the *c*-values. The values obtained for the *a*- and *b*-values were similar, with the *a*-values having a correlation of $r = .85$, and the *b*-values having a correlation of $r = .97$. The *c*-values were less similar, having a correlation of $r = .51$.

The distributions of the item parameter estimates obtained from LOGIST and the scaled ANCILLES estimates are described in Table 6. Although the obtained estimates were highly correlated the statistics shown in Table 6 indicate that there were differences in the item parameter estimate distributions. The *a*-value distributions appear quite similar. However, a dependent *t*-test indicated that the mean ANCILLES *a*-value (.53) was significantly lower than the mean LOGIST *a*-value (.61), yielding a $t = 3.91$ ($p < .01$). A test for the significance of the difference between correlated variances (Ferguson, 1976) yielded a $t = 8.68$, indicating that the variance of the LOGIST *a*-values were significantly greater than the variance of the ANCILLES *a*-values ($p < .01$). Whenever variances were found to be unequal in this study, means were tested for significant differences using the correction in the degrees of freedom set out by Welch (1938). A test for whether the obtained kurtosis values (-.85 for LOGIST, -.72 for ANCILLES) were significantly different from zero (Snedecor and Cochran, 1967) indicated that neither value was significant, as was the case with a test for skewness (Snedecor and Cochran, 1967).

A dependent *t*-test applied to the *b*-value means (-.06 for ANCILLES, -.34 for LOGIST) yielded a $t = 1.97$, indicating that mean ANCILLES *b*-value was greater than the mean LOGIST *b*-value ($p < .05$). A test for the significance of the difference between correlated variances yielded a $t = 6.63$, indicating that the variance of the LOGIST *b*-values ($p < .01$). The greater variance of the LOGIST *b*-values becomes more evident when the range of values is considered. The scaled ANCILLES *b*-values ranged only from -2.88 to 2.34 (a range of 8.34). The kurtosis value for LOGIST (12.21) was significant ($p < .01$), while the kurtosis for ANCILLES (.65) was not. However, the LOGIST *b*-values were significantly negatively skewed ($p < .01$) indicating that, although LOGIST *b*-values go much lower than did ANCILLES, the bulk of the LOGIST *b*-values were actually above the mean of -.34. The ANCILLES *b*-values were not significantly skewed.

Table 3

ANCILLES and LOGIST Item Parameter Estimates

Item No.	ANCILLES			LOGIST		
	a_i	b_i	c_i	a_i	b_i	c_i
1				.15	-1.66	.04
2	.80	1.26	.17	.84	.97	.12
3	.85	1.22	.11	.95	.94	.06
4	.98	.04	.14	1.00	.09	.17
5	.51	-1.92	.06	.29	-2.84	.04
6	.48	-2.13	.06	.26	-3.07	.04
7	.77	.69	.03	.92	.64	.04
8	.69	.17	.15	.47	-.23	.04
9				.08	-7.07	.04
10	.37	-1.37	.03	.25	-1.76	.04
11	.47	.00	.04	.36	-.08	.04
12	.92	.78	.09	.90	.61	.04
13	.57	-.75	.06	.40	-1.11	.04
14	.54	-.21	.01	.44	-.25	.04
15	1.14	-.40	.17	.68	-.90	.04
16	.76	-.27	.02	.61	-.32	.04
17	.58	-.30	.01	.46	-.36	.04
18	.40	-.74	.02	.30	-.89	.04
19	.86	.56	.07	.86	.47	.04
20	.86	.48	.10	.75	.33	.04
21	.55	.16	.16	.36	-.37	.04
22	.50	-.78	.08	.33	-1.28	.04
23	.44	-1.46	.03	.29	-1.92	.04
24	.40	-.24	.02	.31	-.29	.04
25	.99	.62	.22	1.06	.55	.20
26	.89	.25	.09	.74	.11	.04
27	.69	.16	.08	.54	.00	.04
28	.46	-.87	.02	.33	-1.12	.04
29	.51	-.09	.03	.41	-.14	.04
30	.74	.16	.05	.66	.11	.04
31	.64	-.37	.02	.49	-.46	.04
32	.62	.44	.01	.62	.47	.04
33	.97	.81	.05	1.13	.65	.01
34	.54	-.40	.02	.41	-.49	.04
35	.61	.63	.01	.65	.64	.04
36	.85	-.30	.02	.68	-.38	.04
37	.92	.63	.09	.88	.48	.04
38	.87	.90	.02	1.04	.73	.00
39	.57	.55	.07	.49	.44	.04
40	.59	-.16	.01	.48	-.18	.04
41	.53	-.07	.01	.44	-.07	.04
42	.61	-.78	.02	.43	-1.03	.04
43	1.11	.37	.04	1.12	.30	.01
44	.72	1.90	.03	1.17	1.27	.00
45	.85	1.20	.08	1.08	.94	.06
46	.92	.97	.07	1.18	.77	.04
47	.71	.14	.11	.53	-.11	.04
48	.70	.89	.10	.63	.69	.04
49	.53	-.26	.01	.41	-.31	.04
50	.72	-.01	.20	.44	-.62	.04

<u>Note</u>. ANCILLES deleted items 1 and 9 during calibration.

Table 4

ANCILLES and LOGIST Item Parameter
Estimate Correlations

LOGIST	ANCILLES a	ANCILLES b	ANCILLES c
a	.85	.79	.25
b	.56	.97	.14
c	.22	.07	.51

Note: Sample size for both ANCILLES and LOGIST is n = 48.

There were some differences in the distributions of c-values, with the mean ANCILLES c-value significantly higher than the mean LOGIST c-value. However, the actual obtained c-values for the two procedures did not differ greatly in magnitude. For instance, a difference in mean c-values of .02, although significant ($p < .01$), does not seem to be a great difference. The skewness of both distributions (1.14 for ANCILLES, 3.34 for LOGIST) was significant ($p < .01$ for both), but the ANCILLES c-value kurtosis (.60) was not significant, while the kurtosis for LOGIST (13.07) was significant ($p < .01$).

When the item parameter estimates obtained from LOGIST for the two items deleted by ANCILLES are dropped and the comparisons are made only on the 48 items in common, the descriptive statistics change somewhat. The LOGIST mean b-value increases to -.17 without those two items, and the b-value standard deviation drops to .93. The minimum b-value increases to -3.07, the skewness changes to -1.224, and the kurtosis becomes 1.891. Thus, without those two items the b-value distributions from LOGIST and ANCILLES are even more similar. The a-value distributions, however, become slightly less similar when only the 48 common items are considered. The mean a-value for LOGIST becomes .63. This new value slightly increases the difference in the two distributions, as does the new kurtosis value of -.96 and the new skewness value of .56. The new standard deviation (.28) is slightly close to the ANCILLES value, as is the new minimum a-value of .25. The only changes in the LOGIST c-value distribution are to the skewness and kurtosis values, which become 3.26 and 12.42, respectively.

Table 5

ANCILLES Item Parameters Transformed to the LOGIST Parameter Scale

Item No.	a_i	b_i
1	---	---
2	.62	1.52
3	.65	1.47
4	.75	-.06
5	.39	-2.61
6	.37	-2.88
7	.59	.78
8	.53	.11
9	---	---
10	.28	-1.90
11	.36	-.11
12	.71	.90
13	.44	-1.09
14	.42	-.39
15	.88	-.63
16	.58	-.46
17	.45	-.50
18	.31	-1.08
19	.66	.62
20	.66	.51
21	.42	.10
22	.38	-1.13
23	.34	-2.01
24	.31	-.42
25	.76	.69
26	.68	.21
27	.53	.10
28	.35	-1.24
29	.39	-.23
30	.57	.10
31	.49	-.59
32	.48	.46
33	.75	.94
34	.42	-.63
35	.47	.71
36	.65	-.50
37	.71	.71
38	.67	1.06
39	.44	.60
40	.45	-.32
41	.41	-.20
42	.47	-1.13
43	.85	.37
44	.55	2.36
45	.65	1.45
46	.71	1.15
47	.55	.07
48	.54	1.04
49	.41	-.45
50	.55	-.13

Note: The transformation does not alter the c-values.

Table 6

ANCILLES and LOGIST Item Parameter Estimate Descriptive Statistics

Statistic	ANCILLES			LOGIST		
	a_i	b_i	c_i	a_i	b_i	c_i
No. of Items	48	48	48	50	50	50
Mean	.51	-.06	.06	.61	-.34	.04
Median	.53	-.06	.05	.49	-.14	.04
St. Dev.	.15	1.06	.06	.30	1.34	.03
Minimum	.29	-2.88	.01	.08	-7.07	.00
Maximum	.88	2.34	.22	1.18	1.27	.20
Skewness	.36	-.49	1.14	.47	-2.89	3.34
Kurtosis	-.72	.65	.60	-.85	12.21	13.07

Note: Statistics for ANCILLES were obtained using transformed item parameter estimates.

Another analysis performed on the obtained item parameter estimates was to compare the estimates obtained for those items showing lack of fit to the estimates obtained for those items not showing lack of fit. The estimates for which there was significant lack of fit are shown for ANCILLES in Table 7 and LOGIST in Table 8. Examination of these tables does not give any clear indication as to the cause of the lack of fit. The a-values of the items for which there was lack of fit for ANCILLES have a mean not significantly different from the mean of the items not showing lack of fit. The ANCILLES mean b-value for the items showing lack of fit is not significantly lower than the mean b-value for the items not showing lack of fit. For the items for which there was lack of fit for LOGIST the mean a-value is significantly lower than the mean of the a-values for items not showing lack of fit. The mean b-value for the items with lack of fit is not significantly lower than the mean b-value for the poorly fitting items are not significantly different from the c-values for the other items for either procedure. A comparison of the item parameter estimates for the poorly fitting items across the two procedures indicates that there were no significant differences in the means of any of the parameter estimates.

Table 7

ANCILLES Item Parameter Estimates for Items for Which There Was Significant Lack of Fit

Item		a_i	b_i	c_i
2		.62	1.52	.17
5		.39	-2.61	.06
6		.37	-2.88	.06
13		.44	-1.09	.06
16		.58	-.46	.02
17[a]		.45	-.50	.01
18		.31	-1.08	.02
23[a]		.34	-2.01	.03
27[a]		.53	.10	.08
29		.39	-.23	.03
33		.75	.94	.05
42[a]		.47	-1.13	.02
44		.55	2.36	.03
45		.65	1.45	.08
46		.71	1.15	.07
Lack of Fit	$\overline{x}$	.50	-.30	.05
	St. Dev.	.14	1.56	.04
No Lack of Fit	$\overline{x}$	.55	.05	.07
	St. Dev.	.16	.73	.06

[a] Also showed lack of fit for LOGIST.

Table 8

LOGIST Item Parameter Estimates For Items For Which There Was Significant Lack of Fit

Item		a_i	b_i	c_i
17[a]		.46	-.36	.04
22		.33	-1.28	.04
23[a]		.29	-1.92	.04
27[a]		.54	.00	.04
34		.41	-.49	.04
42[a]		.43	-1.03	.04
Lack of Fit	$\bar{x}$	.41	-.85	.04
	St. Dev.	.09	.70	.00
No Lack of Fit	$\bar{x}$	.63	-.27	.05
	St. Dev.	.30	1.40	.04

[a] Also showed lack of fit for ANCILLES.

Ability Estimates

The final set of analyses performed involved the comparison of the ability estimates obtained from LOGIST with the scaled ANCILLES ability estimates. Descriptive statistics for the two obtained ability estimate distributions are presented in Table 9. As can be seen from these statistics the two distributions were quite similar. The range of ability estimates for LOGIST was limited by boundaries of approximately -4.00 to +4.00. In unrestricted operation LOGIST would allow a greater range of ability estimates than would ANCILLES (the same tendency can be noted in the range and variance of b-values).

Table 9

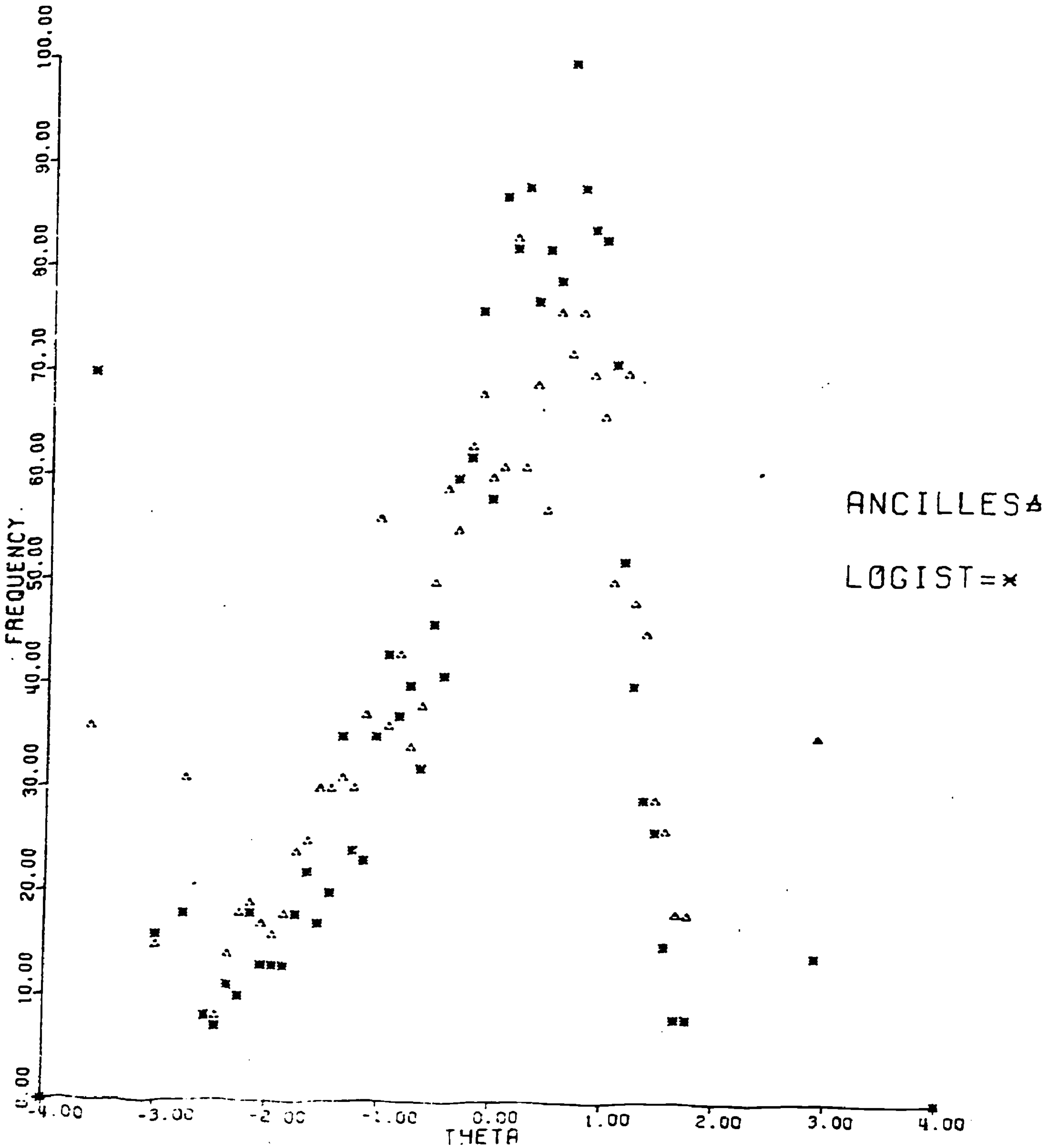
ANCILLES and LOGIST Ability Estimate Descriptive Statistics

Statistics	ANCILLES	LOGIST
No. of Subjects	1999	1999
Mean	-.137	-.137
Median	.045	.142
St. Dev.	1.213	1.214
Minimum	-4.991	-4.061
Maximum	3.303	3.432
Skewness	-.706	-1.164
Kurtosis	.398	1.372

Note: Statistics for ANCILLES were obtained using transformed ability estimates.

The frequency distributions of the ANCILLES and LOGIST ability estimates were plotted together. These frequency distributions are shown in Figure 51. As can be seen in the figure, the two distributions are almost indistinguishable inside the range of -2.00 to +2.00. The only real discrepancy between the two distributions is the height of the LOGIST curve at about -4.00. Because of the arbitrary limits on θ, LOGIST tends to 'pile up' at the limit those examinees whose ability estimates would be outside the limit if the limit were not imposed. This accounts for an unusually large number of ability estimates at approximately -4.00. The great similarity between the two sets of ability estimates is reflected in the correlation of the ability estimates. The Pearson product-moment correlation coefficient obtained for the ability estimates was r = .987. Clearly there is a strong association between the ability estimates assigned by LOGIST and those assigned by ANCILLES.

FIGURE 51
FREQUENCY DISTRIBUTIONS OF
OBTAINED ABILITY ESTIMATES
FOR ANCILLES AND LOGIST



Discussion

When using a Pearson χ^2 statistic to test the goodness of fit of data to a model such as the 3PL model, a number of difficulties are encountered. Before discussing the results of this study, these problems will be addressed and the manner in which they were dealt with in this study will be discussed.

One of the first problems to arise when attempting to compute a chi-square statistic such as was used in this study concerns the formation of intervals on the ability estimate scale. There appears to be some question as to how many intervals to form. For instance, Yen (in press) suggests 10 intervals, while Wright and Mead (1977) recommend six or fewer intervals. The statistic proposed by Wright and Panchapakesan (1969) would require as many intervals as there are obtained number-right scores. Bock (1972), in the fit statistic he has proposed, does not set out any requirements as to the number of categories, but in the example he sets out in his paper (pp. 44-45) he uses 10 intervals. It is clear that the size of the interval will affect the size of the chi-square obtained for the interval. As the interval width increases, the difference between the observed proportions at the ends of the interval and the expected proportion at the center of the interval can be expected to increase. The objective, then, is to have enough intervals (making each interval smaller) to produce sufficiently small within-interval variances in the ability estimates, and thereby reducing within-cell variances of the expected proportions. Alternatively, $\sigma^2_{P_j}$ can be computed and subtracted from the denominator of the chi-square statistic (Wright and Mead, 1977).

In the current study 48 intervals were used. With such a large number of intervals the width of any one interval was sufficiently small as to obviate the need to correct for the variation in expected proportions. However, using such narrow intervals did result in very low frequencies within the extreme intervals, with several intervals having frequencies equal to zero. In order to correct for the small frequencies in the extreme intervals some of the intervals were collapsed together and treated as a single category.

Another problem encountered in applying a chi-square test is the determination of the appropriate degrees of freedom. The degrees of freedom normally associated with the chi-square goodness of fit test when parameters are estimated from the data is

$$df = r - g - 1 \qquad (7)$$

where df is the degrees of freedom, r is the number of categories, and g is the number of parameters estimated from the data (Daniel, 1978). That is, the degrees of freedom are calculated as the number of independent data points (observed proportions) minus the number of independent parameters estimated from the data to produce the expected proportion (Yen, in press). However, when applying the chi-square test to a latent trait model several changes are required. First, because the sum of the expected frequencies is not held fixed, it doesn't really make sense to subtract one from the number of categories. Thus there are r independent data points, rather than

r - 1 (Yen, in press). For the 3PL model there are four independent parameters (θ, a, b, and c) estimated from the data and used in computing the expected proportions. The item characteristic curve for an item is fairly well defined by the computed observed proportions, and the item parameter estimates are clearly dependent on the observed proportions. Therefore, one degree of freedom should be subtracted for each item parameter. However, the ability estimates obtained were dependent upon the entire response vector, and a given item contributes only a small proportion of the information necessary to compute the ability estimates. Therefore, for any given item the estimation of ability entails little loss in degrees of freedom (Yen, in press). Therefore, it is probably more appropriate to subtract g - 1 from the degrees of freedom, rather than g, when using a latent trait model. The degrees of freedom used for this study, then, are given by

$$df = r - (g - 1) \tag{8}$$

where df, r and g are as defined above.

Chi-Square Analyses

It is clear from the results of the chi-square analyses that the LOGIST procedure performed better in terms of goodness of fit. Neither procedure actually fit the test as a whole, but fewer items were rejected when using LOGIST. For the LOGIST procedure only twelve percent of the items showed lack of fit, while for the ANCILLES procedure over thirty percent of the items were rejected for lack of fit.

It is difficult to determine why the lack of fit was significant for ANCILLES more than for LOGIST, especially considering tnat in almost half of the cases (23 out of 48) the LOGIST chi-square was larger than the ANCILLES chi-square. The plots of the expected and observed proportions correct are not very revealing either. However, an examination of the chi-square values obtained for each interval, before being summed, does give some insight as to the cause of the poor fit. For Items 17 and 27 the LOGIST chi-squares were significant due solely to the poor fit in the most positive category, as was the case for Items 16, 17, 18, and 27 for ANCILLES. The last category on the positive end was a very wide category, due to collapsing. Because of this the computed expected proportion, based on the midpoint of the interval, was too high. For Item 27 of ANCILLES, as well as Items 6, 23, 33, 42, 44, and 45, the poor fit was concentrated in the intervals above $\theta = 1.00$. The same was true for Item 23 for LOGIST. For LOGIST, Items 22, 34, and 42 seemed to fit poorly across the ability range, as was the case with ANCILLES for Items 2, 5, 13, and 46. These findings are summarized in Table 10. The poor fit at the extreme ends of the ability range was a problem with both procedures. The poor fit in the most positive interval was a procedural problem, and those items should probably not be counted among those items for which there was significant lack of fit. Without those items there was significant lack of fit for four items for LOGIST and 11 items for ANCILLES.

Table 10

A Summary of the Ranges of Ability for Which Items Showed Poor Fit for ANCILLES and LOGIST

Procedure	Last Interval	Interval where $\theta > +1.0$
ANCILLES	16, 17, 18, 27	6, 23, 27, 33, 42, 44
LOGIST	17, 27	45, 23

MSD Statistic

An examination of the obtained MSD statistics contributes little toward explaining the results. The dependent t-test on these values was significant, which is not consistent with the finding that the ANCILLES chi-squares were not larger than the LOGIST chi-square significantly more than half the time. Moreover, it is disturbing that there was apparently no relationship between the size of the MSD statistics obtained for the items and the size of the chi-square values for the items. A comparison of the MSD values and the item parameter estimates did not yield any clear pattern.

Item Parameter Estimates

A comparison of the item parameter estimates obtained from LOGIST and the transformed ANCILLES estimates also failed to yield a clear explanation. For the full set of items the ANCILLES and LOGIST mean b-values were not significantly different. They were also not significantly different for those items for which there was lack of fit, nor were they significantly different for those items for which there was no lack of fit. For neither procedure was the mean b-value obtained for the items for which there was lack of fit different from the mean b-value for the items for which there was not lack of fit.

The mean a-values for ANCILLES and LOGIST were significantly different. Interestingly enough, however, the mean a-values were not significantly different when considering only those items for which there was lack of fit, nor were they significantly different when considering only the items for which there was not lack of fit. The ANCILLES mean a-value for the items for which there was lack of fit was not significantly different from the mean ANCILLES a-value for the items for which there was not lack of fit. However, for LOGIST the mean a-value for the items for which there was lack of fit was significantly lower than the mean LOGIST a-value for the rest of the items. Because LOGIST yielded higher a-values than ANCILLES for the full

set of items but not for those items for which there was lack of fit it is possible that LOGIST underestimated the a-values for those items for which there was lack of fit. It did appear that LOGIST had more trouble with items with lower discrimination values.

For the full set of items the mean ANCILLES c-value was significantly higher than the LOGSIT mean c-value. The mean ANCILLES c-value for the items for which there was not lack of fit was also greater than the mean LOGIST c-value for the items for which there was not lack of fit. However, when considering only those items for which there was lack of fit, the mean c-values for the two procedures were not significantly different, indicating perhaps that for the items for which there was lack of fit either ANCILLES underestimated the c-values, or LOGIST overestimated the c-value, or both. However, for neither procedure was the mean c-value for the items for which there was lack of fit significantly different from the mean c-value for the rest of the items.

The comparisons of means discussed above do not yield any clear pattern. A comparison of the estimates obtained from ANCILLES and LOGIST with the chi-squares obtained for the procedures does indicate a consistent pattern, however. While it is true that comparing mean values reveals surprisingly few differences in the two sets of item parameter estimates, there is some evidence that the lack of fit of the ANCILLES procedure is related to the item parameter estimates. The correlation of the ANCILLES b-values with the chi-squares obtained for ANCILLES is $r = -.49$. When using the absolute value of the b-values, that correlation is $r = .68$, indicating that the size of the chi-square value obtained for ANCILLES was strongly related to the absolute magnitude of the corresponding b-value. While the mean ANCILLES b-value for the items for which there was lack of fit was not significantly different from the mean for the rest of the items, the variance of the b-values for the items for which there was lack of fit, $s^2 = 2.43$, was significantly higher than the variance of the b-value of the rest of the items, $s^2 = .53$ ($p < .001$). This indicates that the b-values for the items for which there was lack of fit were more extreme than the b-values of the rest of the items. This difference wasn't indicated by the comparison of the means because the extreme values were divided between the positive and negative ends, thus cancelling themselves out when the mean was computed. This pattern does not occur with LOGIST, and the correlation of the LOGIST chi-squares with the absolute values of the LOGIST b-values was $r = 0.0$. It appears, then that at least part of the difference between the fit of the two procedures is accounted for by the poorer ability of ANCILLES to handle extreme b-values.

The correlations of the obtained chi-squares for the two procedures with their respective a- and c-values were not significant. However, a-value estimates also appeared to be a factor in the fit of the LOGIST procedures. For instance, for Item 23 the fit of the model to the data for LOGIST was poorest at the extremes of the ability range. The a-value for Item 23 obtained from LOGIST was $a = .29$, a relatively low discrimination. The a-values for the remaining nonfitting LOGIST items were also low.

Most of the items for which there was poor fit can be accounted for in one of the following ways. For three items for ANCILLES and two items for

LOGIST the poor fit was due to a procedural problem. For the remaining items for LOGIST the poor fit appears to be due to the poor handling of low discrimination values. However, since low discrimination values often indicate multidimensionality, poor fit would be a desired result in these cases (Reckase, 1978). For nine of the remaining 11 items for ANCILLES for which there was lack of fit, the poor fit appeared to be primarily due to the inability of ANCILLES to handle extreme difficulty values. For one of the two remaining items for ANCILLES, Item 29, the poor fit seemed to be across the ability range. Item 29, however, had a low discrimination value, indicating that perhaps ANCILLES also does not handle low discriminators well. For Item 33 the poor fit of ANCILLES was primarily in the intervals where $\theta > +1.0$.

Ability Estimates

As was indicated by Table 9, the ability estimate distributions obtained from ANCILLES and LOGIST were almost identical. Considering the similarity and the fact that the two sets of ability estimates had a correlation of $r = .987$, it is difficult to imagine how the ability estimates could have been a factor in the difference in fit for the two procedures.

Summary and Conclusions

This study was conducted to determine whether there were qualitative differences in the parameter estimates obtained from the ANCILLES and LOGIST estimation procedures. The comparison was made using goodness of fit as a criterion. The results of this study indicate that there are qualitative differences in the estimates obtained from these two procedures. While the parameter estimate distributions obtained from these two procedures were quite similar, lack of fit occurred for significantly more items for ANCILLES than for LOGIST. Further analyses indicated that lack of fit for ANCILLES appeared to be strongly related to item difficulty, while for LOGIST lack of fit was more closely related to item discrimination. It is true that LOGIST is more expensive to use than ANCILLES, but ANCILLES yielded lack of fit significantly more often than LOGIST, and did not yield item parameter estimates for two items. Because of this LOGIST appears to be the procedure of choice.

REFERENCES

Birnbaum, A. Some latent trait models and their use in infering an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of Mental test scores. Reading, MA: Addison-Wesley, 1968.

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Daniel, W. W. Applied nonparametric statistics. Boston: Houghton Mifflin, 1979.

Divgi, D. R. A nonparametric test for comparing goodness of fit in latent trait theory. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April, 1980.

Ferguson, G. A. Statistical analysis in psychology and education. New York, McGraw-Hill.

Marco, G. L. Item characteristic curve solutions to three intractable testing problems. Journal of Educational Measurement, 1977, 14, 139-160.

Reckase, M. D. Ability estimation and item calibration using the one- and three-parameter logistic models: A comparative study (Research Report 77-1). Columbia: University of Missouri, Department of Educational Psychology, November 1977.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 1979, 4(3), 207-230.

Ree, J. M. Estimating item characteristic curves. Applied Psychological Measurement, 1979, 3, 371-385.

Rentz, R. R., and Bashaw, W. L. Equating reading tests with the Rasch model: Final report. Athens, GA: Educational Research Laboratory, University of Georgia, September, 1975.

Snedecor, G. W., and Cochran, W. G. Statistical Methods (6th ed.). Ames, IA: Iowa State University Press, 1967.

Urry, V. W. A Monte Carlo investigation of logistic mental test models. (Doctoral dissertation, Purdue University, 1970). Dissertation Abstracts International, 1971, 31, 6319B. (University Microfilms No. 71-9475).

Urry, V. W. Tailored testing: A spectacular success for latent trait theory. Springfield, VA, National Technical Information Service, 1977. (a)

Urry, V. W. OGIVIA: Item parameter estimation program with normal ogive and logistic three-parameter model options. Washington, D. C.: U. S. Civil Service Commission, Personnel Research and Development Center, 1977. (b)

Urry, V. W. ANCILLES: Item parameter estimation program with normal ogive and logistic three-parameter model options. Washington, D. C.: U. S. Civil Service Commission, Personnel Research and Development Center, 1978.

Welch, B. L. The significance of the differences between two means when the population variances are unequal. Biometrika, 1938, 29, 350-362.

Wood, R. L., Wingersky, M. S., and Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (ETS Research Memorandum RM-76-6). Princeton, NJ: Educational Testing Service, June, 1976.

Wright, B. D., and Mead, R. J. BICAL: Calibrating items and scales with the Rasch model (Research Memorandum No. 23). Chicago: Statistical Laboratory, Department of Education, University of Chicago, 1977.

Wright, B. D., and Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Yen, W. M. Using simulation results to choose a latent trait model. Applied Psychological Measurement, in press.

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

1 Chief of Naval Education and Training
Liason Office
Air Force Human Resource Laboratory
Flying Training Division
WILLIAMS AFB, AZ 85224

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Henry M. Halff
Department of Psychology,C-009
University of California at San Diego
La Jolla, CA 92093

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 CDR Robert S. Kennedy
Head, Human Performance Sciences
Naval Aerospace Medical Research Lab
Box 29407
New Orleans, LA 70189

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Gro
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Scien
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for th
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Dexter Fletcher
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria,VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RES INSTITUTE
5001 EISENH NUE
ALEXANDRIA, 3

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
FT Benjamin Harrison, IN 46256

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Dr. Marty Rockway
Technical Director
AFHRL(OT)
Williams AFB, AZ 58224

1 Jack A. Thorp, Maj., USAF
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary
of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

5

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Rm T-255
Educational Testing Service
Princeton, NJ 08450

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charl Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2