SUMMARY COVER SHEET

1. Title: Introduction of the Computer in the Statistics Curriculum Submitted to the Office of Computing Activities Under the Program for Developing Computer Uses.

2. Name of Institution Hope College

Address: City Holland State Michigan Zip Code 49423

Grant to: Hope College
(Official name of institution to which grant should be made)

3. Clarence Handlogten, Treasurer and Business Manager
(Name and title of chief administrative officer of institution)

4. Elliot A. Tanis — Director's name
481-34-8023 — Social Security No.
Mathematics — Department

Telephone: 616 — Area Code
Ext. 2266 — Office No.
392-5111 — Dept. No.
296-8384 — Home Phone

5. Period covered in this proposal: June, 1971 (Starting date) to August, 1973 (Ending Date)

6. Total amount requested from the National Science Foundation $45,800

7. Elliot A. Tanis — Signature of proposed director
Clarence Handlogten — Signature of authorizing official

8. September 18, 1970 — Date of submission
Clarence Handlogten, Treasurer & Business Manager — Name and title of official authorized to sign for institution

PROPOSAL: INTRODUCTION OF THE COMPUTER IN THE STATISTICS CURRICULUM

Elliot A. Tanis
Herbert L. Dershem
Hope College
Holland, Michigan
September 1, 1970

The modern digital computer has greatly enhanced the teaching of statistics to undergraduates. Many studies have been made and courses introduced which use the computer as a tool in the teaching of statistics (see references). These efforts have been directed along five general lines.

1. Generating empirical data. Whereas previously students would spend many hours throwing dice, flipping coins or drawing cards, the computer can provide a much larger volume of data in a matter of seconds.

2. Illustrating statistical principles. Statistical principles that were only illustrated by textbook examples can now be "discovered" by the student by generating the empirical data and then testing the validity of the principle under study.

3. Stored programs. The use of special purpose subroutines to perform a complicated statistical procedure (such as analysis of variance or regression analysis) gives the student a chance to input numbers and receive an analysis from which he may gain experience in interpreting results.

4. Storing raw data. The computer can be used to store a large amount of raw data collected from real life observations in such a way that students and instructors can perform statistical analyses of this data.

5. Computer-assisted instruction. A system which allows for

interaction between the student and the computer can be utilized by allowing the machine to tutor the student, i.e., to take him step by step through the process of learning the statistical principles, spending as much time at each step as the student needs.

Each of these uses of the computer in the statistics curriculum has great merit when employed in the proper situation. We wish to examine the effect of introducing the computer into our statistics curriculum at Hope College in the above ways.

We feel, however, that the computer can serve as a more valuable tool when the student must do a significant amount of the programming himself. If a student is able to program the correct solution of a problem, he must have a more than superficial understanding of how the problem is solved. He must take into account all possible situations that might be encountered and all possible sets of input data. This proposal is made with the purpose of incorporating this idea into the statistics curriculum at Hope College.

DESCRIPTION OF PROPOSED COURSES

The Hope College Mathematics department currently offers two course sequences in statistics: Math 35, Introductory Statistics and and Math 55-56 Mathematical Probability and Statistics. We propose to modify these courses by introducing the computer into the course material in an integral way. Our method for doing this and the rationale follow for the two sequences considered separately.

At the present time there are two courses offered by the mathematics department at Hope College which are either required or strongly recommended for majors in the fields of psychology, economics and business administration, sociology and political science. These two courses are Mathematics 27, Introduction to Digital Computers and

Mathematics 35, Introductory Statistics. The former is a two semester - hour course in which the students learn to program in Fortran on the IBM 1130 while the latter is a three semester - hour course giving a general introduction to statistics. The catalog description of these two courses is found in Appendix A.

We propose to combine these two one-semester courses into a two-semester sequence, three hours each semester. It is felt that each course could benefit from the material of the other. Computer programming can be taught more readily when the students are motivated by interesting problems from their major area which require statistical analysis. By learning programming in conjunction with the statistics necessary for analyzing these problems, the students could understand the analysis behind the programs they write and gain practice in interpreting their results. This would be a significant improvement over the present arrangement where programming must be taught without assuming any statistical background for the student and hence statistical programs assigned are usually nothing more than formulas to be translated to Fortran. For examples of the types of programs related to statistics currently assigned in the Math 27 course, see Appendix B.

The above could be presented as an argument for making elementary statistics a prerequisite for the programming. But we also believe that an argument for the opposite approach can also be given leaving their combination as an optimal solution. We again emphasize that this course would differ from most currently established computer-aided statistics courses inasmuch as we would have students writing their own programs rather than depending on previously prepared statistical routines. (We do not, however, rule out the use of such routines in this course where we feel they would be useful). The computer can be an effective pedagogical tool since the student must thoroughly understand a technique before he can write a program to perform it. This fact has been long utilized in numerical analysis, but seldom in connection with statistics.

The proposed course would not be simply a combination of the two present courses, but would involve an interweaving of the material in such a way that the student always has learned enough programming to program the next statistical technique and knows enough statistics that the latest programming principle learned can be illustrated by examples involving previously learned statistics. This obviously will require a large amount of planning. The programming material will necessarily be taught in a sequence which differs from the normal practice. For example, subscripted variables will be introduced early in the course rather than toward the end as is usual since most statistical applications require their use. Seldom used programming concepts will be placed near the end of the course.

Since the proposed course is a total of six semester-hours in length and the present courses total five, there will be time for the introduction of additional topics in statistics and computer science with which social and behavioral scientists should be familiar and which are not included in the current curriculum. Some such topics are random number generators, Monte Carlo techniques and simulation. The students would also be expected to complete a project. This would involve beginning with raw data from a well-defined experiment or observation, and analyzing and refining this data statistically in order to draw certain conclusions from the data. The class would be divided into teams of students with similar interests and each team would attack a problem related to their field.

The effectiveness of this course will depend substantially on the quality of the problems and data used for these projects as well as problems used for classroom examples and homework assignments. We will obtain these with the assistance of our colleagues in the various departments from their own research and from textbooks and research articles in their areas of interest. In this way we hope to obtain a collection of interesting, useful and non-trivial problems. A useful side effect for the professor providing the problem is that the class analysis might provide some additional insight for him in his own work.

MATHEMATICAL PROBABILITY AND STATISTICS

We teach a two semester, three hours per semester, junior-senior level mathematical probability and statistics course (see Appendix A). For this course we believe it is important for the students to learn the theory of probability and statistical inference.

We further believe that it is important to illustrate the theory. This can often be done very simply with dice, cards or a spinner. However, for many examples the computer is an invaluable aid.

To illustrate the Central Limit Theorem, it is instructive to sample from several distributions, using different sample sizes. In this way the student gets some feeling for the fact that

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

has a limiting N(0,1) distribution. See Appendix C for an example of this idea.

For Monte Carlo techniques and simulation, a pseudo-random number generator is necessary. Given a deterministic method for generating "random numbers", it is an instructive exercise for students to test statistical properties of the numbers generated. E.g., we can find the probability density functions of order statistics. Now generate 8000 sets of 4 random numbers. Order each set of 4 numbers. Do the 8000 smallest numbers represent a random sample from the distribution of the smallest order statistic? See Appendix C for an expansion of these ideas.

Confidence intervals for means can be simulated to illustrate, e.g., 90% confidence intervals. See Appendix C.

If $X_1, X_2, ---, X_n$ is a random sample of size n from a normal distribution, mean μ, variance σ^2, then $\sum_{i=1}^{n} (X_i - \mu)^2/\sigma^2$ has a chi-square distribution with $\gamma = n$ degrees of freedom while $\sum_{i=i}^{n} (X_i - \overline{X})^2/\sigma^2$ has a chi-square distribution with $\gamma = n-1$ degrees of freedom. This loss of one degree of freedom can be illustrated through simulation on the computer.

In order for simulation on the computer to most effectively illustrate the theory, it would be beneficial for each student to write most of his own programs. This would necessitate making programming a prerequisite for this course. Selected subroutines should be made available, e.g. for ordering data or grouping data. However, that part of the program which illustrates the theory in probability and statistics should be written by the student in order that he derive the most benefit from this experience.

The fact that this procedure is effective has been demonstrated by 5 Hope College students who did write their own programs for undergraduate research projects.

In order that there would be sufficient course time to cover all of the necessary material, we propose to add a two hour laboratory each semester. It is in developing materials for this laboratory that we desire the support of NSF.

In addition to illustrating theory, we would like to gather data sets of live data in the same way as proposed for the introductory statistics course. It will be possible in many cases for the two classes to share the same data. In this way students can learn to apply statistical inference in real-life situations.

PREVIOUS EXPERIENCE WITH COMPUTERS IN STATISTICS

Much previous work has been done in introducing the computer into the statistics curriculum here at Hope. Dr. Jay E. Folkert was a participant in the University of North Carolina NSF project on computers and statistics. Dr. Folkert experimented with the use of the computer mainly in the form of previously prepared subroutines to illustrate the theory in the introductory statistics course. Dr. Herbert Dershem did the programming and provided consultation for this project.

Dr. Elliot Tanis has used the computer extensively in his mathematical probability and statistics course and in undergraduate research projects in statistics (see Appendix D). It is from this experience, the recommendations of Dr. Folkert and Dr. Tanis and the comments of their students that the proposed revisions to our statistics curriculum arise.

PROPOSAL

We propose to accomplish the modifications to the statistics curriculum at Hope College in a five phase program over a period stretching from the Summer of 1971 to the Summer of 1973. This period includes two academic years and will therefore afford the opportunity to teach both course sequences twice.

Phase I of the project will occur in the Summer of 1971. During this period Dr. Tanis and Dr. Dershem will make preparations for the courses to be taught the following academic year. This will involve the planning and arranging of material in a logical sequence for the introductory statistics and programming course, the development of meaningful experiments and useful real-life data sets, the design of useful problems and examples for the computer, and the writing of certain routines. The two professors will be assisted by four students, all of whom will have taken one year of mathematical statistics. These students will be trained to be lab assistants the following academic year.

The second phase will span the 1971-72 academic year and will consist of the teaching of the two proposed course sequences for the first time. Dr. Dershem will teach the introductory statistics and programming course while Dr. Tanis will teach the mathematical probability and statistics course. Each will consult with the other throughout the year. Students will assist with the laboratory. It is anticipated that it will be necessary to distribute notes in each of these courses as there will be no textbooks which follow our proposed course requirements. The four students who were involved the previous summer will now serve as laboratory assistants. Their duties will include assistance in the preparation of notes, maintaining software developed in the previous summer, develop new software as the need arises, assist students in program debugging and operate the computer for the students. One quarter release time is proposed for Drs. Dershem and Tanis for the preparation of notes.

In the Summer of 1972, phase III of the project is planned. This will involve the development of new software, the designing of new experiments and the holding of a conference. This conference will be for educators from the Midwest who are interested in this program. A detailed description of this conference is found in the section "Dissemination of Information". Two students who completed the mathematical probability and statistics sequence in the 1971-72 academic year will be employed as assistants for Phase III.

Phase IV will include the academic year 1972-73 and consist of the second teaching of both of the new courses. New approaches will be tried, replacing old ineffective approaches and new examples, projects and experiments will be attempted. Revisions will be made to the notes prepared a year before. Four students who have completed the new mathematical probability and statistics sequence will be employed to assist as in the previous year. Two of these students will have assisted in Phase III.

The Summer of 1973 will be the fifth phase. In this phase the entire program will be evaluated and recommendations made for future years. Those experiments judged the most effective and our final conclusions regarding the project will be accumulated and prepared for distribution to all who attended the conference in Phase III and any other interested educators.

DISSEMINATION OF INFORMATION

In the summer of 1972 a conference will be scheduled to report the results of our first year of study and to solicit suggestions for improvement. Invited to this conference will be interested members of the mathematics departments of all four year colleges in the Midwest area who have computer facilities. We will limit the attendance at this conference to approximately fifty participants. The conference will be scheduled in August to allow adequate preparation and so that it is after the completion of most summer sessions.

The conference will begin on Monday with an afternoon session, followed by morning and afternoon sessions Tuesday through Thursday and a final session on Friday morning. All sessions will be chaired by Dr. Tanis or Dr. Dershem. Copies of interesting problems and outlines of the courses as they stand at this time will be distributed to the participants.

The following summer, an update of these materials will be made and sent to all participants of the conference and to anyone else who has requested to be added to our mailing list.

PERSONNEL

The principal personnel will be Dr. Elliot Tanis and Dr. Herbert Dershem. Dr. Tanis will have the primary responsibility for the mathematical probability and statistics course while Dr. Dershem will direct the development of the introductory statistics and programming course. A resume of these two professors is found in Appendix E.

Dr. Jay Folkert, chairman of the department of mathematics at Hope College, will serve as a consultant. Dr. Folkert was a participant in the NSF - University of North Carolina project on computers in introductory statistics and taught the introductory statistics course at Hope College using the computer for one year.

FACILITIES

The principal facility is the IBM 1130. Hope College has a full complement of supporting equipment for the 1130 and an extensive calculator center complete with programmable calculators to handle smaller computations easily. For a complete listing of facilities currently available, see Appendix F.

Of the six 029 keypunches owned by the college, four are available for student use. It is expected that the introduction of these two new courses will greatly increase the demand for these already overtaxed keypunches. Hence, funds for renting two additional keypunches are included in the budget of this proposal.

The computer center will be able to supply more than an adequate amount of computation power to handle the needs of this project. Educational programs of students and professors have top priority on the machine and turn around time for student jobs is seldom more than one hour at any time. Many students find it more convenient to run their programs in the evening. The computer center is currently kept open in the evenings for this sole purpose. One of the purposes of the student assistants in this project is to allow the time available for student runs in the evening to be expanded by having the assistants operate the computer for an additional portion of the evening hours in order to minimize the impact of these new courses on the normal day-time computer schedule.

BUDGET

Phase I - Summer, 1971

Salaries:

1)	Elliot Tanis, 2/9 of $13,500	$3000.00
2)	Herbert Dershem, 2/9 of $12,600	2800.00
3)	4 student assistants, 12 weeks @$60/wk	2880.00

Other direct costs:

4)	Keypunch rental, 3 months @$69.30/month	207.90
5)	Reference material	100.00
6)	Consulting fees	500.00
7)	Expendable equipment and supplies-2% of other direct operating costs	190.00

Phase I direct costs - total $9677.90

Phase II - Academic Year, 1971-72

Salaries:

1)	Elliot Tanis, 1/4 of $13,500	$3375.00
2)	Herbert Dershem, 1/4 of $12,600	3150.00
3)	4 student assistants, 5 hr/wk, $1.50/hr. for 30 weeks	900.00

Other direct costs:

4)	Keypunch rental, 2 for 9 months @$69.30/month	1247.40
5)	Expendable equipment and supplies - 2% of other direct operating costs	174.00

Phase II direct costs - total $8846.40

Phase III - Summer, 1972

Salaries:

1)-2)	Same as Phase I	$5800.00
3)	Two student assistants, 12 weeks @$60.00/wk	1440.00

Conference expenses:

4)	Travel expenses for conference participants, 50 @500 miles/participant @8¢/mile	$2000.00
5)	Room and board, 4 days for 50 participants @$10.00/day	2000.00
6)	Publicity	200.00

Other direct costs:

7)	Keypunch rental, 2 for 3 months @ $69.30/month	415.80
8)	Expendable equipment and supplies - 2% of other direct operating costs	237.00

Phase III direct costs - total $12092.80

Phase IV, Academic Year 1972-73

Salaries:

1)	4 student assistants, 5 hr/wk for 30 wks	$ 900.00

Other direct costs:

2)	Keypunch rental, same as 4) Phase II	1247.40
3)	Expendable equipment and supplies - 2% of other direct operating costs	43.00
		$2190.40

Phase V, Summer, 1973

Salaries:

1)	Elliot Tanis, 1/9 of $13,500	$1500.00
2)	Herbert Dershem, 1/9 of $12,600	1400.00
3)	Secretarial	500.00

Other direct costs:

4)	Expendable equipment and supplies - 2% of other direct operating costs	68.00

Phase V direct costs - total $3468.00

Total budget - All five phases

1) Salaries - staff	$21,025.00
2) Salaries - students	6,120.00
3) Keypunch rental	3,118.50
4) Reference material	100.00
5) Consulting fees	500.00
6) Conference expenses	4,200.00
7) Expendable equipment and supplies	712.00
8) Secretarial	500.00
Total direct costs	$36,275.50
Indirect costs - 45.3% of 1)	9,524.32
Total costs - direct and indirect	$45,799.82

Supplemental Budget Information

Salaries. The salaries of Dr. Tanis and Dr. Dershem are based on estimations of their average nine month salaries over the period of funding. These amounts will vary slightly from the actual amounts budgeted, but they should be very close to correct.

Consulting fees. The consulting fees budgeted will be used to reimburse the members of the social and behavioral science faculty who serve as consultants by providing the project personnel with data and problems for use in the proposed courses.

Expendable equipment and supplies. The expendable equipment and supplies, computed at 2% of all other direct operating costs, will include computer supplies, such as printer forms, IBM cards and plotter paper, duplicating supplies and stationery.

Indirect costs. The indirect costs were computed in accord with the negotiated agreement dated April 28, 1970. A xerox copy of this agreement is contained in Appendix H.

Computer time. The charge for use of the Hope College IBM 1130 is $30.00 per hour. All computer charges related to this project will be a contribution to the project by Hope College. It is estimated that this project will require over $9,000 worth of computer time.

APPENDIX A

CATALOG DESCRIPTION OF CURRENT COURSES

Mathematics 27. Introduction to Digital Computers - Capabilities of digital computers and their place in society. Computer languages with emphasis on FORTRAN. Laboratory experience with a computer including programming of a major problem in the student's field of interest. Two Hours.

Mathematics 35. Introductory Statistics - A general introduction to the area of statistics for students majoring in other departments. Includes study of the binomial and normal distributions with applications of estimation and testing of hypotheses, non-parametric methods, regression and correlation.

Mathematics 55. Mathematical Probability and Statistics I - Concepts of probability, probability and distribution function of random variables, mathematical expectation, variance, and moment generating functions of distributions. Prerequisite: Calculus II (may be taken concurrently).

Mathematics 56. Mathematical Probability and Statistics II - Continuation of Mathematics 55 emphasizing statistics. Sampling theory, estimation, testing of hypothese, regression and correlation, introduction to analysis of variance. Prerequisite: Mathematics 55.

APPENDIX B

PROGRAMS RELATED TO STATISTICS ASSIGNED IN MATH 27, 1969-70

1. Write a program to process a deck of cards; compute the mean and standard deviation and print the count, mean and standard deviation.

2. Write a program that reads a value of N from a card and then reads N integer values, all between 0 and 9 inclusive, punched 80 to a card. The program is to print the mode (or modes) of this set of numbers and how many times the mode was attained.

3. Write a program that reads the number of observations N from a card followed by N cards, each containing a value of X and the corresponding value of Y and computes the regression coefficients, the standard error of estimate and the coefficient of correlation for these data sets.

4. Write a subroutine which accepts as inputs an integer N, N real values A(1), ..., A(N), an integer M, and M boundary values for a collection of frequency classes, B(1), ..., B(M). It is assumed that the values of B are in ascending order. The subroutine is to return to the calling program the mean and standard deviation of the values A, for each frequency class, the number of data values in that class, the percentage of data values in that class, and the cumulative percentage of data values up to and including that class.

APPENDIX C

THE CENTRAL LIMIT THEOREM AND MIXED DISTRIBUTIONS

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution which has mean, μ, and a finite variance, σ^2. Let $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$. The Central Limit Theorem states that

$$W_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has a limiting normal distribution with mean, 0, and variance, 1. (Let N(0,1) denote this standard normal distribution.)

When n is "sufficiently large," W_n has an approximate N(0,1) distribution. A "rule of thumb" that is often used is that when n is larger than about 25, W_n will appear to have a N(0,1) distribution.

Note that the hypotheses of the Central Limit Theorem say very little about the distribution from which the sample is taken. In particular the Central Limit Theorem is applicable to mixed distributions.

We shall give an example which illustrates the conclusion of the Central Limit Theorem empirically when the sample is taken from a mixed distribution. The simulation for this example was done on Hope College's IBM 1130 computer.

EXAMPLE Roll a fair die. If the outcome is even, let X equal this observed value. If the outcome is odd, spin a spinner which selects a number at random from the interval [0,1]. Let X equal the number selected. The random variable, X, defined by the "game" has a mixed distribution which has a probability of 1/6 at each of the integers 2,4, and 6, and has a uniform density of 1/2 on the interval [0,1].

The distribution function of X is defined by

$$F(x) = \begin{cases} 0, & x < 0, \\ x/2, & 0 \le x < 1, \\ 1/2, & 1 \le x < 2, \\ 4/6, & 2 \le x < 4 \\ 5/6, & 4 \le x < 6, \\ 1, & 6 \le x. \end{cases}$$

The mean and variance of X are defined by

$$\mu = \int_{-\infty}^{\infty} x \; dF(x)$$

$$= \int_{0}^{1} x/2 \; dx + 2(1/6) + 4(1/6) + 6(1/6)$$

$$= 9/4;$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 \, dF(x) - (9/4)^2$$

$$= \int_0^1 x^2/2 \, dx + 4(1/6) + 16(1/6) + 36(1/6) - (81/16)$$

$$= 71/16.$$

Let $X_1, X_2, \ldots, X_n$ denote n plays of the "game", i.e., a random sample of size n from the distribution for which F is the distribution function. For several values of n we shall use a chi-square goodness of fit test to test the hypothesis, H, that

$$W_n = \frac{\overline{X} - 9/4}{\sqrt{(71/16)/n}}$$

has a $N(0,1)$ distribution.

Let the points -2.55, -2.25, -1.95, ..., 2.25, 2.55 partition the set of real numbers into 19 intervals. The end intervals are infinite rays and the remaining 17 intervals are of equal length, .30. Let p_i equal the probability that W_n falls in the i'th interval when H is true. E.g., $p_2 = P(-2.55 < W_n \leq -2.25)$, when W_n has a $N(0,1)$ distribution.

We shall generate m = 1000 w_n's on the computer where each w_n is based on n plays of the "game." Let Y_i denote the observed number of w_n's in the i'th interval. Then

$$\chi^2 = \sum_{i=1}^{19} \frac{(Y_i - 1000p_i)^2}{1000p_i}$$

has an approximate chi-square distribution with $\nu = 18$

degrees of freedom when H is true.

Instead of setting a significance level for this test, for each value of n between 2 and 15, inclusive, we ran three independent chi-square goodness of fit tests, each based on a sample of size m = 1000 w_n's. In table 1 we have recorded the calculated values of χ^2. For this chi-square value we have also recorded the (1-p)'th quantile, $\chi^2(p)$. I.e., if U has a chi-square distribution with ν degrees of freedom, then

$$P(U \geq \chi^2(p)) = p.$$

When ε is recorded, $\chi^2(p) < .0005$.

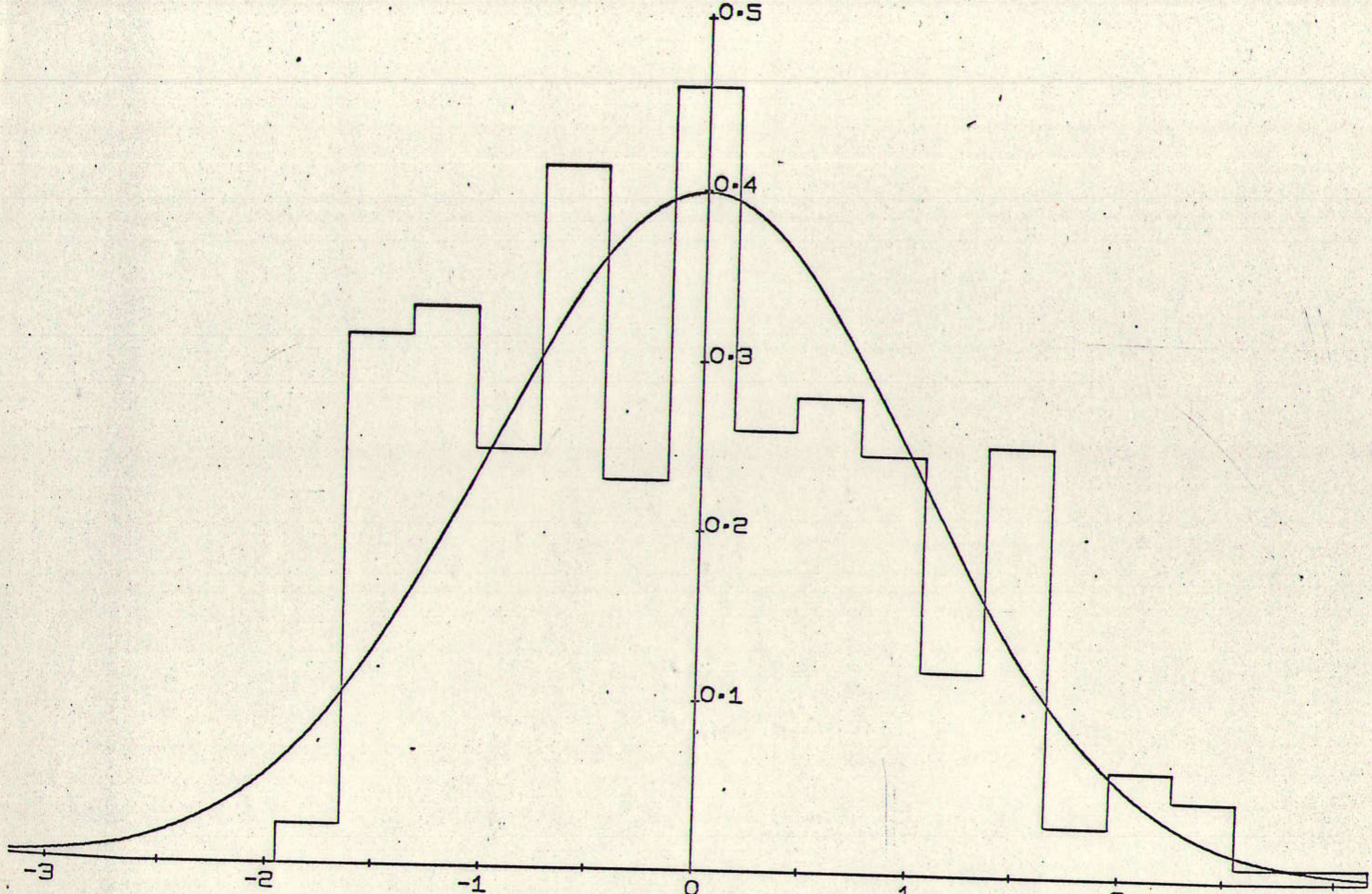
For this mixed distribution, W_n appears to have a N(0,1) distribution when n is larger than about 8 or 9.

In order to show graphically the convergence of W_n to the N(0,1) distribution, in figures 1,2,3, and 4 we have graphed the probability density function for the N(0,1) distribution along with the relative frequency histogram for each set of data which yielded the starred chi-square values in table 1.

n	χ^2	$\chi^2(p)$	χ^2	$\chi^2(p)$	χ^2	$\chi^2(p)$
2	753.20	ε	639.07	ε	614.34	ε
3	247.60*	ε	160.22	ε	189.49	ε
4	76.72	ε	100.44	ε	106.72	ε
5	48.76	ε	52.85	ε	48.61	ε
6	23.91*	.158	48.17	ε	35.18	.009
7	30.50	.033	32.22	.021	47.76	ε
8	26.11	.097	29.66	.041	24.09	.152
9	11.70	.862	27.98	.062	20.09*	.328
10	15.48	.629	22.43	.213	18.11	.449
11	25.43	.114	29.68	.041	25.53	.111
12	17.65	.479	13.23*	.778	27.96	.063
13	22.75	.200	13.01	.791	13.83	.740
14	12.06	.844	23.25	.181	34.02	.013
15	13.17	.781	19.88	.340	24.38	.143

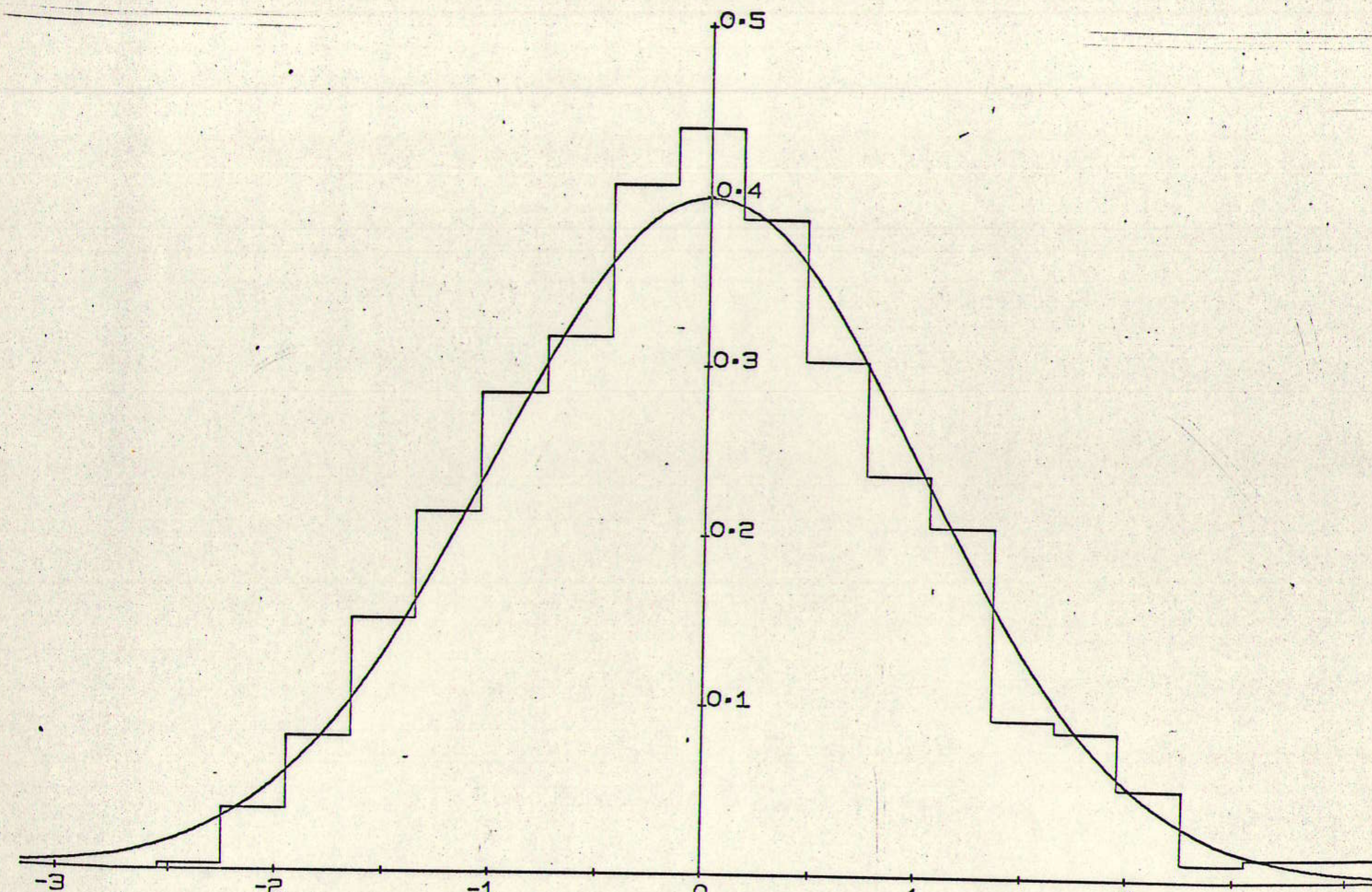
Table 1

* See figures 1,2,3, and 4.



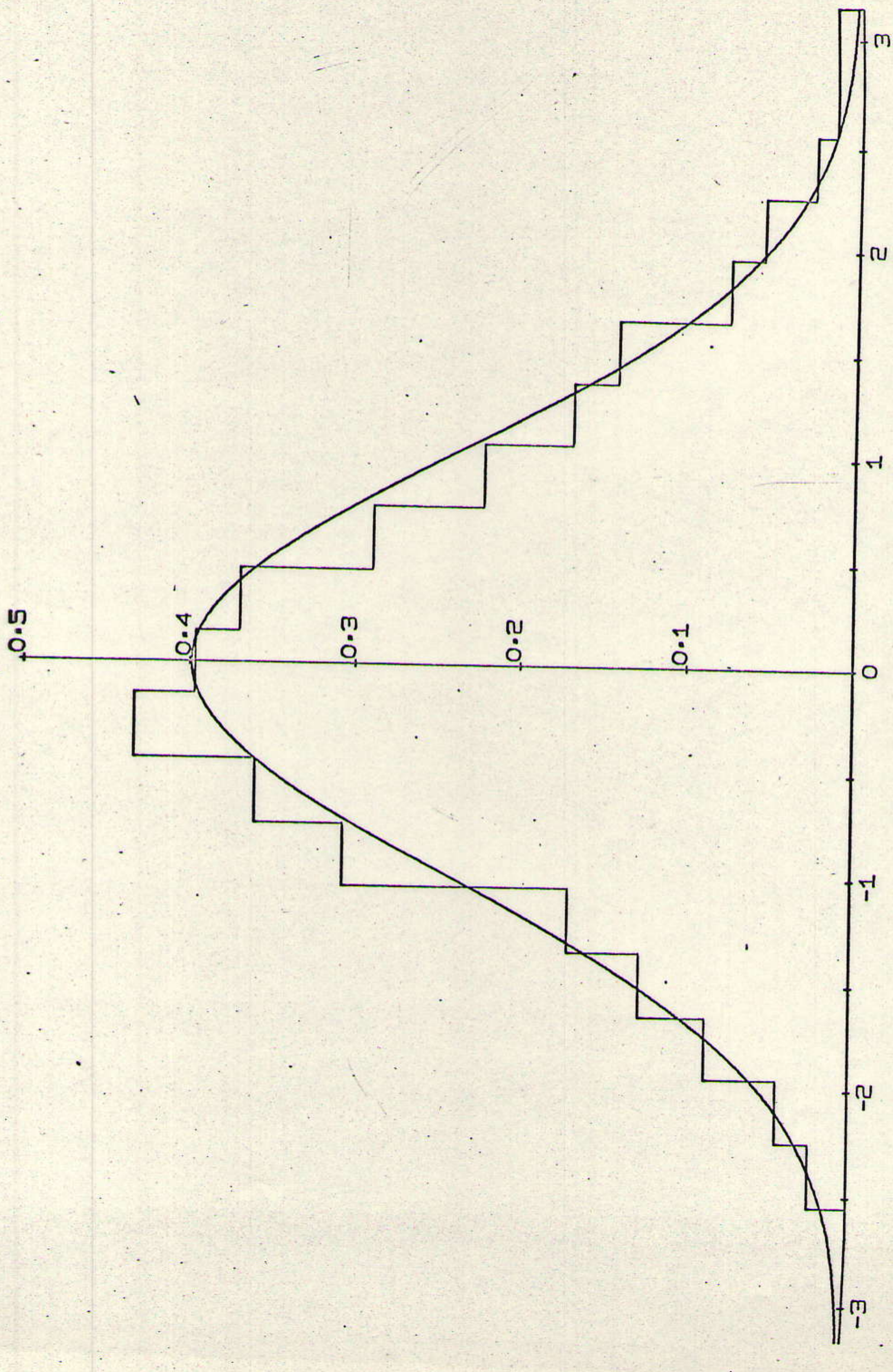
N = 3, CHI-SQUARE = 247·60

FIGURE 1



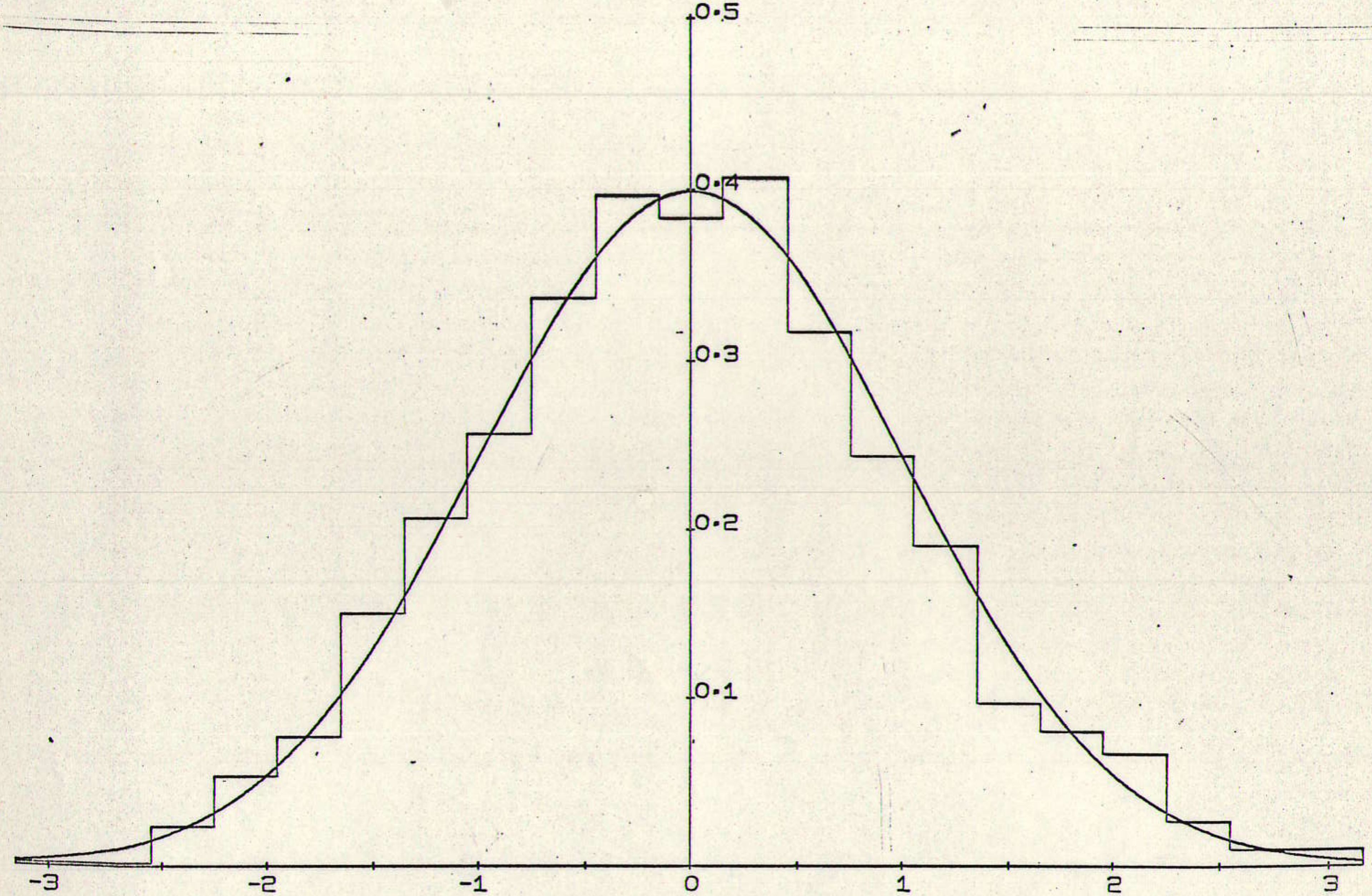
N = 6, CHI-SQUARE = 23·91

FIGURE 2



N = 9, CHI-SQUARE = 20·09

FIGURE 3



N = 12, CHI-SQUARE = 13·23

FIGURE 4

ORDER STATISTICS

Let Y_1, Y_2, Y_3, Y_4 be the order statistics of a random sample of size 4 from the uniform distribution on the interval [0,1]. I.e., if x_1, x_2, x_3, x_4 are the observed sample values, then $y_1 = \min\{x_1, x_2, x_3, x_4\}$, ..., $y_4 = \max\{x_1, x_2, x_3, x_4\}$ are the observed sample values of the order statistics.

The probability density functions of the 4 order statistics are given by

$$
\begin{aligned}
g_1\ (y_1) &= 4(1-y_1)^3, & 0 \le y_1 \le 1,\\
g_2\ (y_2) &= 12y_2(1-y_2)^2, & 0 \le y_2 \le 1,\\
g_3\ (y_3) &= 12y_3^2(1-y_3), & 0 \le y_3 \le 1,\\
g_4\ (y_4) &= 4y_4^3 &, \quad 0 \le y_4 \le 1.
\end{aligned}
$$

Each of these functions is equal to zero elsewhere.

The probability density function of the range, $R = Y_4 - Y_1$, is given by

$$h(r) = 12r^2(1-r)\ ,\ 0 \le r \le 1,$$

zero elsewhere.

Eight thousand sets of four numbers were generated, each set representing a random sample of size 4 from the uniform distribution on the interval [0,1]. Each set of 4 numbers was ordered from the smallest to the largest. Using the 8000 sets of ordered numbers, five chi-square goodness of fit tests were run, four tests for the order statistics and one test for the range. The interval [0,1] was partitioned into subintervals of length .05 for each of the five tests. The theoretical probabilities for the subintervals were found using the respective probability density functions. A good fit was accepted for each test at an α = .05 significance level.

In figures 5,6,7,8, and 9 are graphed the probability density functions and relative frequency histograms for Y_1, Y_2, Y_3, Y_4, and R, respectively.

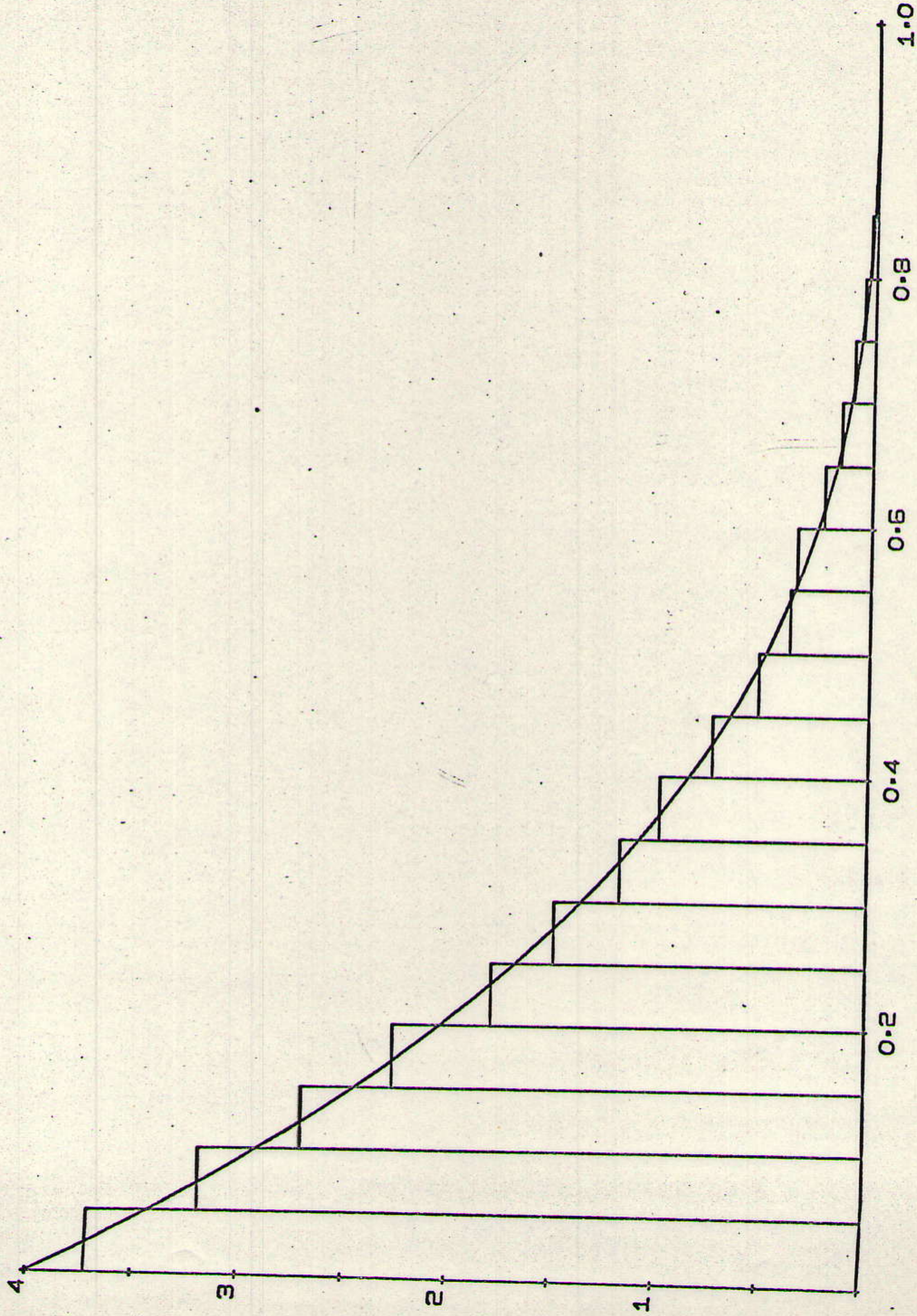


FIGURE 5

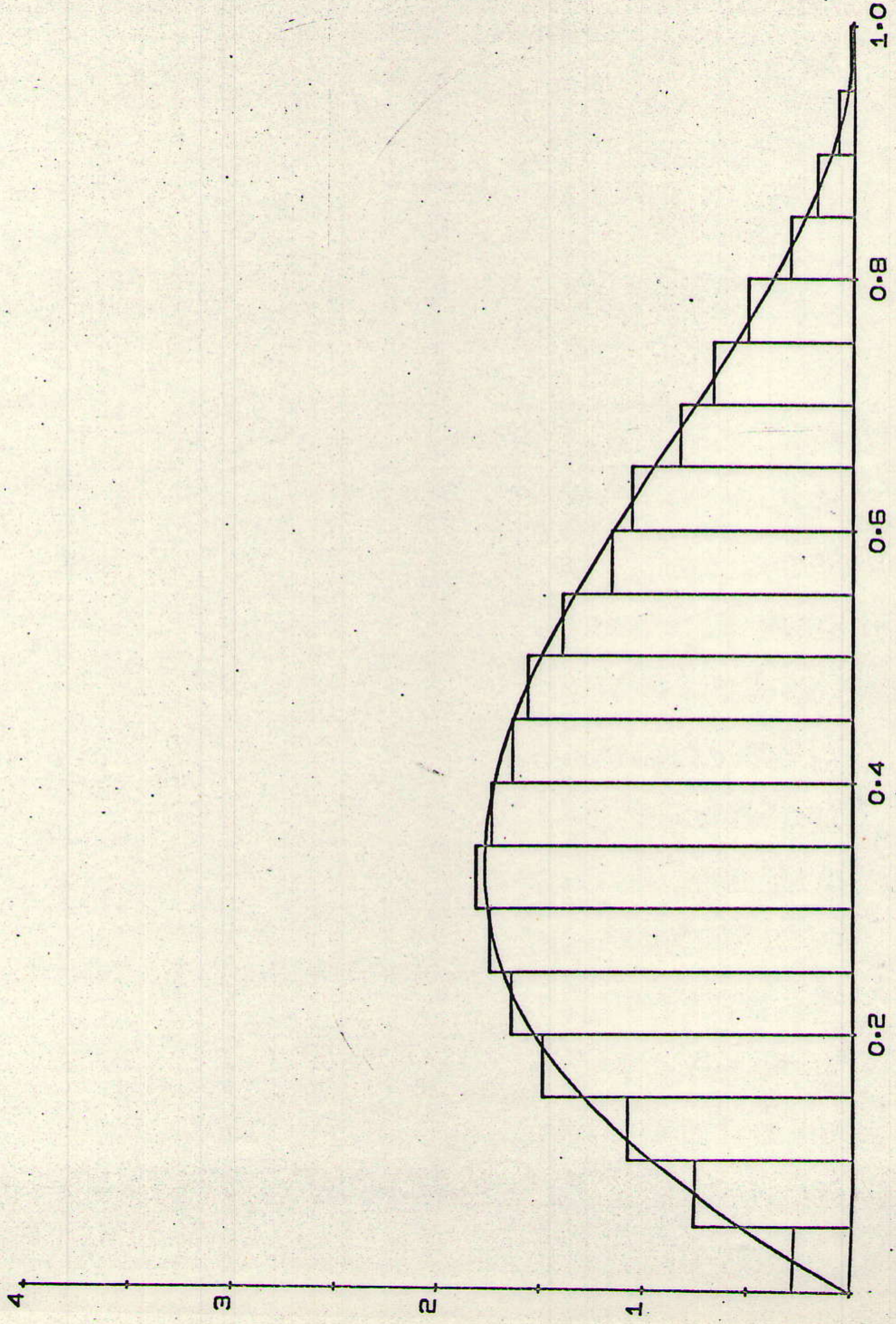


FIGURE 6

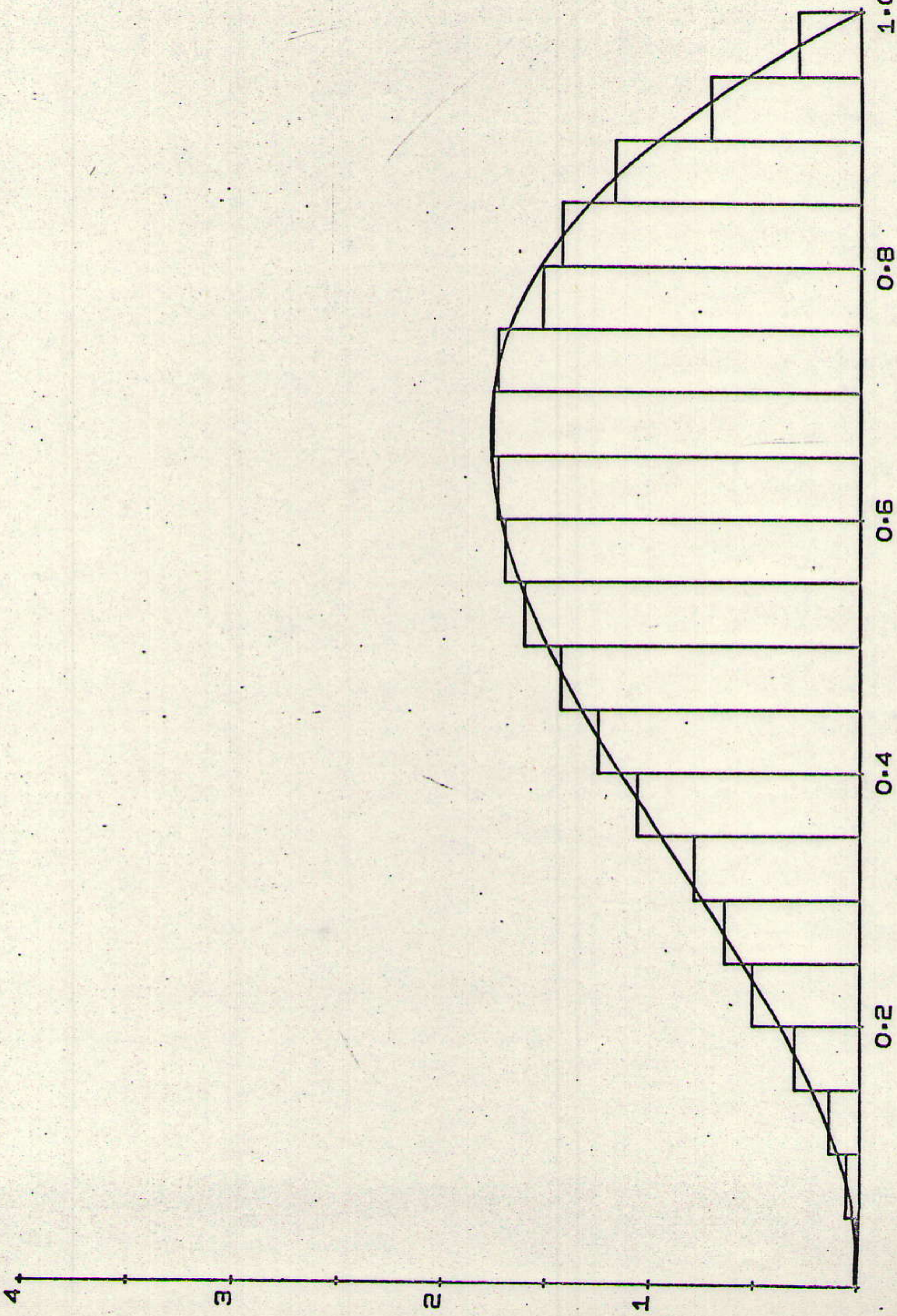


FIGURE 7

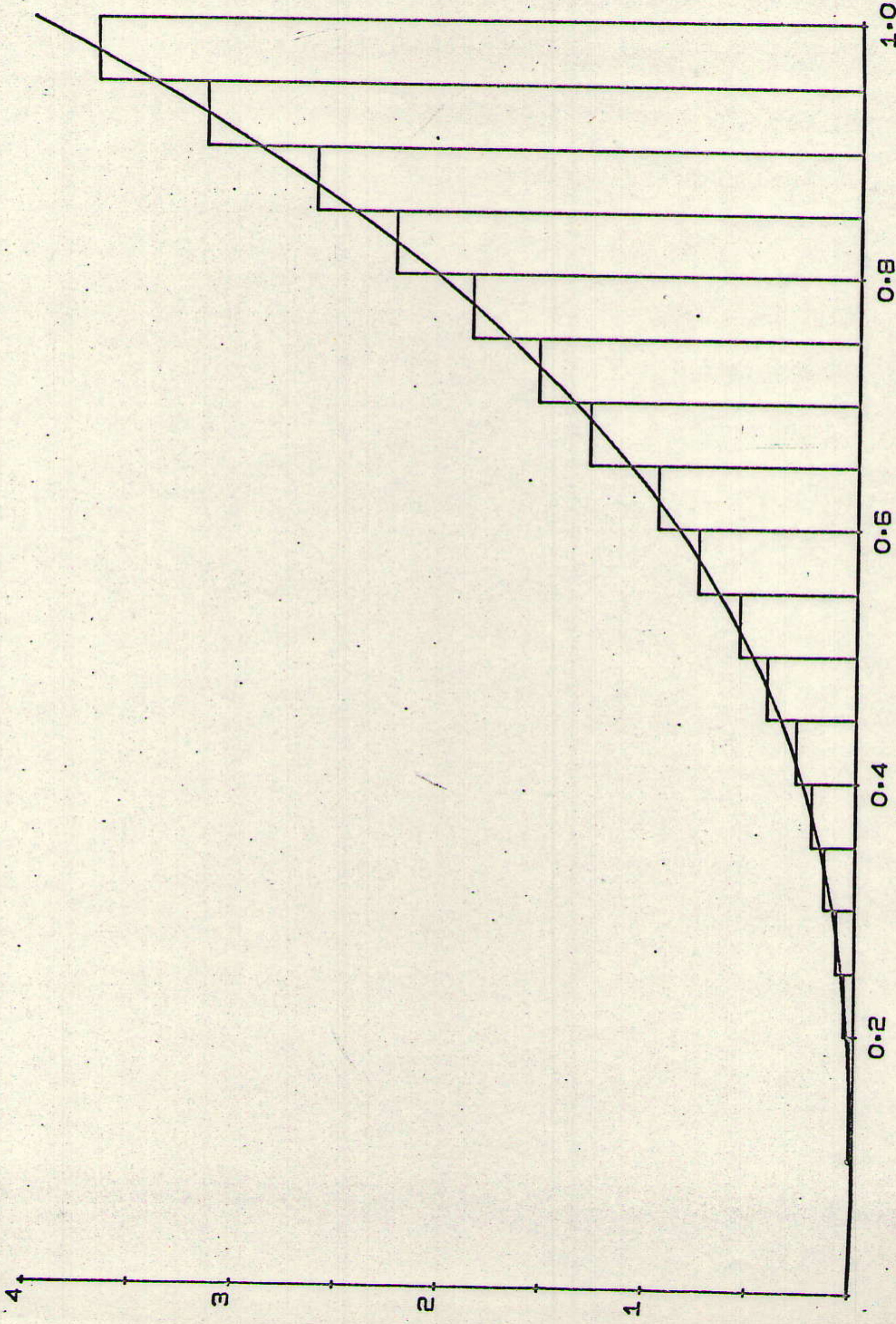


FIGURE 8

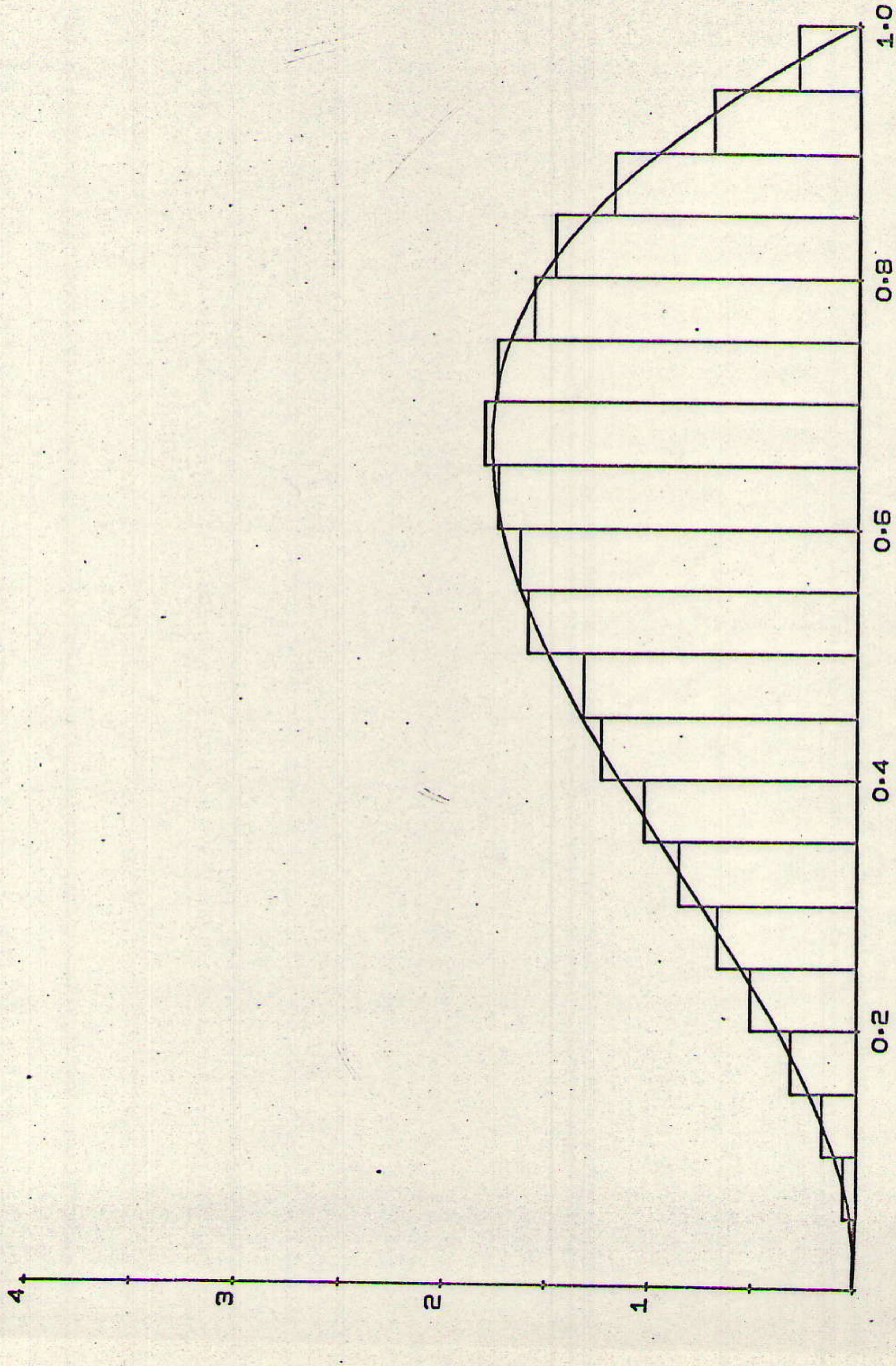


FIGURE 9

CONFIDENCE INTERVALS

Let x_1, x_2,...,x_{10} be the values of a random sample of size 10 from a N(μ,12.34) distribution. A 90% confidence interval for the unknown mean, μ, is

$$[\bar{x} - 1.645\sqrt{12.34/10},\ \bar{x} + 1.645\sqrt{12.34/10}].$$

For a particular sample this interval either does, or does not, contain the mean, μ. However, if many such intervals were calculated, it should be true that about 90% of these intervals contain the mean, μ.

Fifty two random samples of size 10 from a N(40,12.34) distribution were simulated on the computer. For each of these 52 random samples, a 90% confidence interval was calculated as if the mean, μ = 40, were unknown. In figure 10 each of these 52 intervals is represented by a horizontal line.

For the 52 calculated intervals, 46 or 88.5% contain the mean.

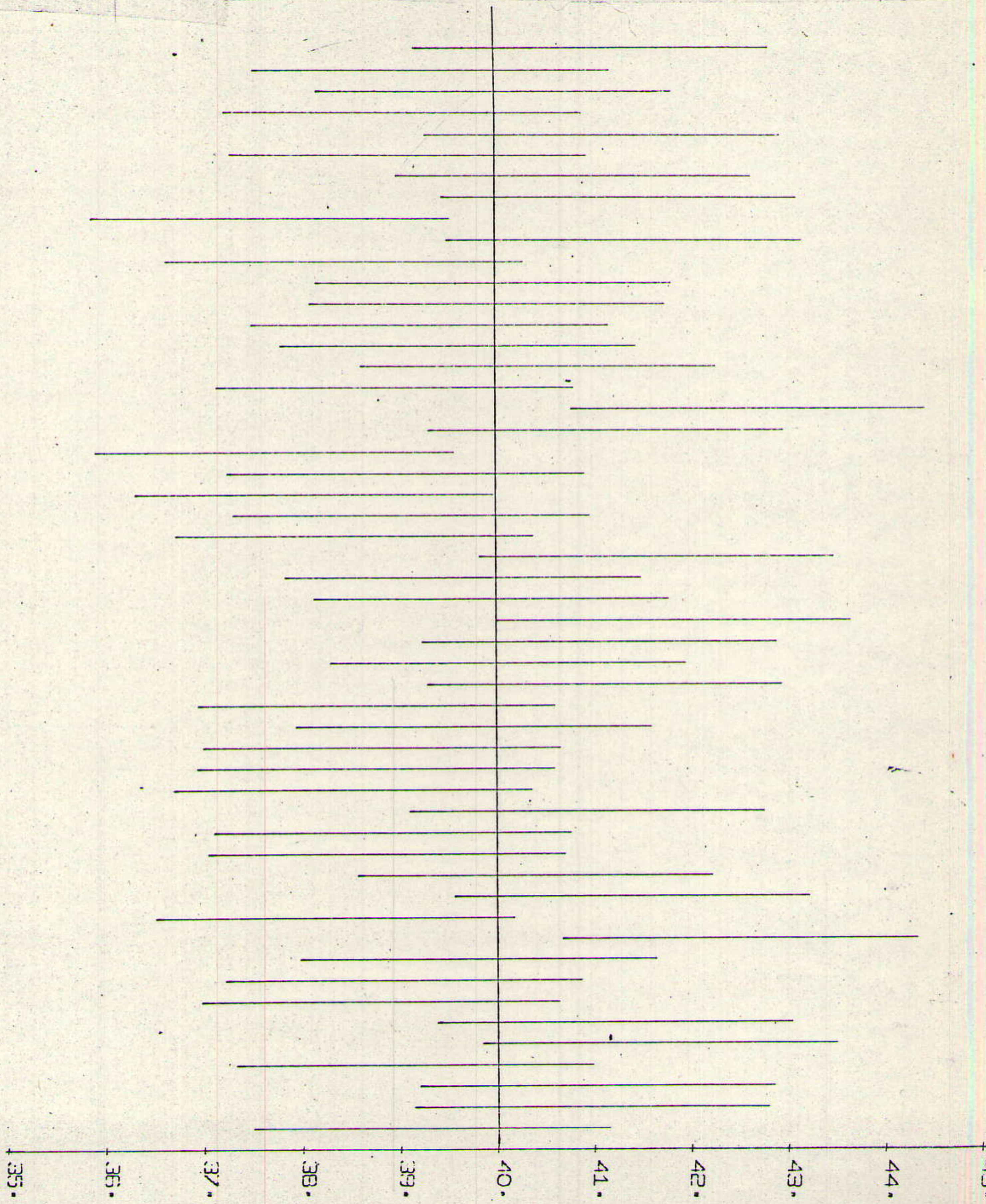


Figure 10

APPENDIX D

Research and independent study projects in Statistics at Hope College during the last 5 years.

1965-1966: Wepfer, Richard - "Criteria in Statistical Decision Theory": The problem may be stated: A statistician must decide which action is to be selected from a set of actions in order to do himself the most good in a situation in which nature may be in any one of several possible states. The paper presents the minimax and Bayesian decision criteria and compares their usefulness.

1967-1968: Gross, Deanna - "How Large is Large Enough? A Study of the Ramifications of the Central Limit Theorem Through Simulation Techniques": The paper illustrates the validity and power of the Central Limit Theorem by means of Monte Carlo techniques with the IBM 1130 computer. It shows that for different distributions different sample sizes are necessary to justify a normal approximation. (This paper was presented to the annual 1969 meeting of the Michigan Chapter of the Mathematical Association of America. In addition part of the results have been submitted to The American Statistician for publication.)

1968-1969: Diggelmann, Henry Wm. - "Random Number Generation and Chi-square Testing": The paper discussed pseudo-random number generators applicable on the IBM 1130 computer. (An article based on this phase of the project has been submitted to Communications of the Association for Computing Machinery.) Another phase of the project considered ramifications of the rule of thumb for chi-square goodness of fit tests which states that the expected number of outcomes in each cell should be at least 5.

Lang, Patricia - "Application of the Central Limit Theorem to a Mixed Distribution": Let $\overline{X}$ be the sample mean of a random sample of size n from a mixed distribution having mean, μ, and variance, σ^2. For two particular mixed distributions, this paper illustrates empirically that $W_n = \sqrt{n}(\overline{X}-\mu)/\sigma$ has an approximate normal distribution, mean 0, and variance 1, when n is at least equal to 10. (Part of the results of this paper have been submitted to The American Statistician. This work was conducted as a result of questions raised in a mathematical statistics class. No extra academic credit was given.)

Mehnert, Elizabeth F. - "Applications of Central Limit Theorem to Discrete Distributions and Existence of Means": The paper considers the size of n necessary for $W_n = \sqrt{n}(\overline{X}-\mu)/\sigma$ to have an approximate normal distribution, mean 0, variance 1, where $\overline{X}$ is the sample mean of a random sample from a discrete distribution with mean, μ, and variance, σ^2. The second topic considered was an interpretation empirically of what it means to say that a mean and/or variance of a distribution does not exist.

1970(summer): Newell, Joyce - "Effects of Round-off on the t-test": In order to apply a t-test, the population from which the sample is taken is assumed to be normal. In general, live data has been rounded off so that in effect the sample has been taken from a discrete population. The effect of round off, as a function of the population variance and the sample size, was studied using Monte Carlo techniques on the IBM 1130 computer.

APPENDIX E

RESUMES

Dr. Elliot A. Tanis

Education:

B. A.	Central College, Pella, Iowa, 1956.
M. S.	University of Iowa, 1960 (Mathematics)
Ph.D.	University of Iowa, 1963 (Mathematics) Thesis: "Linear Forms in the Order Statistics from an Exponential Distribution"

Teaching Experience:

Teaching Assistant, University of Iowa, 1959-63.
Assistant Professor of Mathematics, University of Nebraska, 1963-65.
Associate Professor of Mathematics, Hope College, 1965 - Present.

Other Experience:

U. S. Army, 1957 - 1959.
Associate Director, NSF Summer Institute for Advanced Placement Teachers of Mathematics, Hope College, 1967, 1968.
Director, NSF Summer Institute for Advanced Placement Teachers of Mathematics, Hope College, 1969.
Lecturer, NSF Summer Institute for Advanced Placement Teachers of Mathematics, Hope College, 1968, 1970.

Professional Affiliations:

Member of:
- Mathematical Association of America
- American Mathematical Society
- Institute of Mathematical Statistics
- American Statistical Association
- Society of Sigma Xi

Publications:

"An Iterated Procedure for Testing Equality of Several Exponential Distributions", American Statistical Journal, Volume 58, 1963. (with Dr. Robert V. Hogg)

"Linear Forms in the Order Statistics from an Exponential Distribution," The Annals of Mathematical Statistics, Volume 35, 1964.

"A Random Number Generator Which Uses the Sum of Two Congruential Generators," submitted to Communications of the ACM.

"The Central Limit Theorem and Mixed Distributions," submitted to The American Statistician.

"A Card Matching Problem," submitted to the Mathematical Log.

"A Statistical Hypothesis Test for the Classroom," submitted to The Mathematics Teacher.

Dr. Herbert L. Dershem

Education:

B. S.	University of Dayton, Dayton, Ohio, 1965
M. S.	Purdue University, 1967 (Computer Science)
Ph.D.	Purdue University, 1969 (Computer Science)
	Thesis: "Approximation of Bessel's Differential Operator of Fractional Order by Finite Difference Operators"

Teaching Experience:

Teaching Assistant, Purdue University, 1968-69.
Assistant Professor of Mathematics, Hope College, 1969 - Present.

Other Experience:

Part time computer programmer, University of Dayton Research Institute, 1961-1965.
Numerical analyst, University of Dayton Research Institute, Summers of 1965-67.

Professional Affiliations:

Member of:

American Mathematical Society
Association for Computing Machinery

Publications:

"An Algorithm for Approximating the Solution to a General First Order System of Differential Equations," UDRI-TM-67-137, University of Dayton Research Institute, 1967.

"Approximation of the Bessel Eigenvalue Problem by Finite-Differences," submitted to SIAM J. Numer. Anal., 1970.

"Bessel Difference Systems of Fractional Order," submitted to Math. Anal. and Appl., 1970.

APPENDIX F

Hope College Computer, Punch Card and Calculator Equipment

6	IBM 029 Card Punches
1	IBM 59 Card Verifier
1	IBM 83 Card Sorter
1	IBM 519 Reproducing Punch
1	IBM 85 Collator
1	Standard Register Series 1400 Burster
1	IBM 1131 Central Processing Unit
1	IBM 1133 Multiplex Control Enclosure
1	IBM 1442 Card Read Punch
1	IBM 1403 Printer
1	IBM 2310 Disk Storage Unit
1	Calcomp 565 Plotter
15	IME 86 Electronic Desk Calculators
6	IME 308 Digicorders (Programmers)
1	IBM Typewriter Printout

APPENDIX G

HOPE COLLEGE

Hope College is a four year, undergraduate liberal arts institution granting the degrees of bachelor of arts and bachelor of music. Hope is fully accredited by The North Central Association of Colleges and Secondary Schools, and The National Council for Accreditation of Teachers Education. The current enrollment of Hope is approximately 2,000 students.

The mathematics department of Hope College consists of nine members, four of whom hold the Ph.D. degree and two who are completing their dissertation for the doctorate. The department has graduated 94 majors in the last 5 years - 28 of these entered graduate school and 5 entered professional school. There are currently 44 junior-senior mathematics majors enrolled.

APPENDIX H

Xerox Copy of Indirect Cost Negotiation Agreement Between Hope College and the Department of Health, Education and Welfare.

NEGOTIATION AGREEMENT

INSTITUTION: Hope College
Holland, Michigan 49423

DATE April 28, 1970

FILING REF.: This replaces Negotiation Agreement dated January 5, 1970

SUBJECT: Indirect cost rate(s) for use on grants and contracts with the Department of Health, Education, and Welfare.

SECTION I: This Section shall be used for funding indirect costs on current grants and contracts. The rate(s) cited below was based on an indirect cost proposal dated submitted in a timely manner for FYE August 31, 1969.

Type	Effective Period From	To	Rate*	Locations	Applicable To
Provisional	9/1/69	Until Amended	45.3%	On-Campus	All Programs
Provisional	9/1/69	Until Amended	19.2%	Off-Campus	All Programs

*Base: Direct salaries and wages including vacation, holiday and sick pay, but excluding other fringe benefits.

Treatment of fringe benefits: Fringe benefits applicable to direct salaries and wages are treated as direct costs.

SECTION II: This Section shall be used for processing indirect cost rates on public vouchers and reports of expenditures.

Type	Effective Period From	To	Rate*	Locations	Applicable To
Final	9/1/67	8/31/68	44.03%	On-Campus	All Programs
Final	9/1/67	8/31/68	21.79%	Off-Campus	All Programs
Final	9/1/68	8/31/69	45.3%	On-Campus	All Programs
Final	9/1/68	8/31/69	19.2%	Off-Campus	All Programs
Provisional	9/1/69	Until Amended	Use rates, etc. cited in Section I above.		

*Base: Same as cited in Section I above.

Treatment of fringe benefits: Same as cited in Section I above.

For purposes of determining indirect costs to be applied to grant reports of expenditures, utilizing rates cited in this Section, the grantee elects Alternative 1 cited in Section III.

DGAP SF 1
5/29/69

White

SECTION III: Rate Alternatives

The following rate alternatives are for use on grants only, and are not applicable to contracts. The application of the rates in Section II to contracts will be governed by the provisions of each contract. The alternative cited in Section II together with the applicable rate cited in Section II shall be applied consistently to all DHEW grants.

Alternative 1. The rate established for the fiscal period in which the grant budget year begins will be applied to the entire grant budget year.

Alternative 2. A single rate will be applied to the entire grant budget year and the rate that will be used is the one that is effective for the longest period during that budget year. If the budget year falls equally in two rate periods, the rate of the earlier period will be used.

Alternative 3. The rates established for the periods in which direct expenditures are actually made will be applied to those expenditures.

SECTION IV: General

A. LIMITATIONS: Use of the rate(s) contained in this agreement is subject to any applicable statutory or administrative limitations. Acceptance of the rate(s) agreed to herein is predicated upon the conditions: (1) that no costs other than those incurred by the grantee/contractor were included in its indirect cost rate proposal and that such costs are legal obligations of the grantee/contractor, (2) that the same costs that have been treated as indirect costs have not been claimed as direct costs, and (3) that similar types of costs have been accorded consistent treatment.

B. ACCOUNTING CHANGES: If a predetermined rate(s) is contained in this agreement it is based on the accounting system in effect at the time the proposal was submitted. Changes to the method of accounting, including but not limited to, changes in charging a particular type of cost from indirect to direct, require the prior approval of the Director of Grants Administration Policy, DHEW. Failure to obtain such approval may result in subsequent audit disallowances.

C. NOTIFICATION TO FEDERAL AGENCIES: Copies of this document may be provided to other Federal offices as a means of notifying them of the agreement contained herein.

D. SPECIAL REMARKS: None

ACCEPTANCE:

Grantee: Hope College

By: C. J. Vanderlugt
Name

Treasurer
Title

May 11, 1970
Date

Donald V. [illegible]
for Henry G. Kirschenmann, Chief
Cost Policy & Rate Negotiation Branch

Date May 18, 1970

Negotiated by B. Feldman

Telephone 202 962-8334

SF 1

REFERENCES

Appelbaum, Mark I. and Guthrie, Donald. "Use of Computers in Undergraduate Statistics Instruction," _Proceedings of a Conference on Computers in the Undergraduate Curricula_, The University of Iowa, 1970, pp. 2.1-2.3.

Cooley, William W. "Computer-Assisted Instruction in Statistics," _Statistical Computation_, Academic Press, New York, 1969.

Garrett, Henry G. "SAMDS, A Program to Generate Empirical Sampling Distributions of the Mean," _Proceedings of a Conference on Computers in the Undergraduate Curricula_, The University of Iowa, 1970, pp. 2.32-2.38.

Gilgen, Albert R. and Hall, Michael A. "Computer-Assisted Teaching of Experimental Design," _Proceedings of a Conference on Computers in the Undergraduate Curricula_, The University of Iowa, 1970, pp. 2.39-2.46.

Koh, Young O. "The Computer as an Instructional Tool for the Statistics Course," _Proceedings of a Conference on Computers in the Undergraduate Curricula_, The University of Iowa, 1970, pp. 2.4-2.17.

Lohnes, Paul R. and Cooley, William W. _Introduction to Statistical Procedures: with Computer Exercises_, John Wiley & Sons, Inc., New York, 1968.

Sterling, Theodor D. and Pollack, Seymour V. "Use of the Computer to Teach Introductory Statistics," _Communication of the Association for Computing Machinery_, _9_, 4 (1966), pp. 274-279.

Sterling, Theodor D. and Pollack, Seymour V. _Introduction to Statistical Data Processing_. Prentice-Hall, Englewood-Cliffs, New Jersey, 1968.

Wallace, David L. "Computers in the Teaching of Statistics: Where Are the Main Effects?" Statistical Computation, Academic Press, New York, 1969.

Wikoff, Richard L. "Using the Computer in Basic Statistics Courses," Proceedings of a Conference on Computers in the Undergraduate Curricula, The University of Iowa, 1970, pp. 2.18-2.24.

Technical Summary Report

Name of Institution:	Hope College
Address:	Holland, Michigan 49423
Principal Investigators:	Elliot A. Tanis Herbert L. Dershem
Grant Number:	GJ - 28786
Period covered by grant:	June, 1971 to August, 1973
Title:	Introduction of the Computer in the Statistics Curriculum

SUMMARY

The purpose of this grant is twofold: (i) Materials are being deveolped for an introductory statistics course in such a way that the student always has learned enough programming to program the next statistics technique and knows enough statistics so that the latest programming principles learned can be illustrated by examples involving previously learned statistics.

(ii) Materials are being developed for a laboratory for a two semester junior-senior level mathematical probability and statistics course in which students have the opportunity of increasing their understanding and appreciation of the theory by performing simulations on the computer.

During the summer of 1971 Professors Dershem and Tanis, along with four capable students wrote exercises and developed supporting software for our IBM 1130 computer.

During the 1971-72 academic year the course changes were introduced into our curriculum. The reaction of the students to both courses was favorable.

Materials for both courses are being revised during the summer of 1972. By the end of this summer they will be available for limited use in other colleges and universities. In particular these materials will be shared with approximately 50 educators from other institutions at a Conference in August which is supported by this grant.

Personnel

The following professors are associated with and supported by this grant:

Dr. Herbert L. Dershem, Assistant Professor of Mathematics, Hope College.

Dr. Elliot A. Tanis, Professor of Mathematics, Hope College.

The following students received support during the summer of 1971 and the 1971-72 academic year:

Lynn Klaasen Hillegonds

Timothy Hillegonds

Richard Pohl

Glenn Weener

The following students are receiving support during the summer of 1972:

Roger Crisman

Richard Meyers

Gail Ringsmith

Papers

The following papers were presented at the 1972 Conference on Computers in Undergraduate Curricula and appear in the *Proceedings* of that Conference.

"A Course on Computing and Statistics for Social Science Students", by Herbert L. Dershem

"Theory of Probability and Statistics Illustrated by the Computer", by Elliot A. Tanis

Conference

A Summer Conference for college teachers of mathematics on the use of the computer in the undergraduate statistics curriculum will be held August 14-18, 1972. This conference is being supported by this grant. During this conference we will share our experiences.

Elliot A. Tanis
Elliot A. Tanis

June 27, 1972
Date

ANNUAL PROGRESS REPORT
July 1, 1972-June 30, 1973

Name of Institution: Address:	Hope College Holland, Michigan 49423
Principal Investigators:	Elliot A. Tanis Herbert L. Dershem
Grant Number:	GJ - 28786
Period Covered by Grant:	June, 1971 to August, 1973
Title:	Introduction of the Computer in the Statistics Curriculum

1) HIGHLIGHTS OF PROJECT ACTIVITY

A conference held on the Hope College campus on August 14-18, 1972 was attended by 52 college teachers of statistics from throughout the Midwest. At this conference the philosophy and the materials of the two courses developed under this grant were presented to the participants in formal sessions. During more informal sessions, the participants were invited to react to the material and make suggestions for improvement. Copies of all material developed, including exercises, program descriptions and program printouts, were given to all participants. In addition, materials were sent to over 20 other educators throughout the country by special request.

During the 1972-73 academic year, the two statistics courses for which material is being developed under this grant were taught for the second time by Professors Tanis and Dershem using the modifications made during the preceding summer. In addition, part of the materials were used by many of the participants in the Summer Conference.

A report of this project was made by Dr. Dershem at the annual meeting of the Michigan Section of the Mathematical Association of America on May 4, 1973.

A report of this project was made by Dr. Tanis at the Fourth Conference on Computers in the Undergraduate Curricula at the Claremont Colleges on June 20, 1973.

A Senior Honors Project, which made extensive use of the computer software developed in this project, was completed by Roger Crisman, one of our mathematics majors.

A final revision of materials and preparation for dissemination is being carried out during the summer of 1973. In addition, reactions are being collected from those educators who used the material this past year.

2) <u>PUBLIC INFORMATION RELEASE</u>

In order to effectively use a computer as an aid in education, it is necessary to develop both course materials and programs for the computer. For educators to develop these materials, they need time in addition to ability. This grant from the National Science Foundation provided the time for professors and students at Hope College to work together so that future students at Hope and other institutions will receive a better education.

In particular a course was developed for teaching computer programming and statistics in an integrated way to students in the social sciences. In addition a laboratory, which is now an integral part of a mathematical probability and statistics course, was developed which uses the computer to illustrate empirically many theoretical results.

3) <u>PERSONNEL</u>

The following professors are associated with and supported by this grant:

Dr. Herbert L. Dershem, Associate Professor of Mathematics, Hope College.

Dr. Elliot A. Tanis, Professor of Mathematics, Hope College.

The following students received support during the summer of 1971 and the 1971-72 academic year:

Lynn Klaasen Hillegonds

Timothy Hillegonds

Richard Pohl

Glen Weener

The following students received support during the summer of 1972 and the 1972-73 academic year:

Roger Crisman

Richard Meyers

The following student received support during the summer of 1972:

Gail Ringsmith Buis

The following professor is receiving support during the summer of 1973:

Professor John T. Whittle, Assistant Professor of Mathematics, Hope College.

4) <u>PUBLICATIONS</u>

Tanis, Elliot A., "Theory of Probability and Statistics Illustrated by the Computer," <u>Proceedings of The 1972 Conference on Computers in Undergraduate Curricula</u>, Atlanta, Georgia, 1972, pp. 513-520.

Dershem, Herbert L., "A Course on Computing and Statistics for Social Science Students," Proceedings of The 1972 Conference on Computers in Undergraduate Curricula, Atlanta, Georgia, 1972, pp. 525-528.

Tanis, Elliot A., "A Computer Laboratory for Mathematical Probability and Statistics," Proceedings of a Fourth Conference on Computers in the Undergraduate Curricula, Claremont, California, 1973, pp. 416-426.

Elliot A. Tanis
Elliot A. Tanis

June 27, 1973
Date

A Course on Computing and Statistics for Social Science Students

Herbert L. Dershem
Department of Mathematics
Hope College
Holland, Michigan 49423
616 392-5111

Origin of the course. Before the introduction of the course to be described in this paper, most economics and social science students at Hope College enrolled in two mathematics courses. These were introductory statistics and computer programming. In the academic year 1969-70 two developments occurred which pointed to the advisability of combining these courses.

First, Dr. Jay Folkert conducted experimental sections of the introductory statistics course at Hope College in which he used the computer as a tool to obtain illustrative information. He did this by allowing students to use previously written programs on prepared data decks. This was done in conjuction with the project to study the use of the computer in statistical instruction sponsored by the National Science Foundation and the University of North Carolina, in which Dr. Folkert was a participant. At the close of the experiment it was Dr. Folkert's feeling that the computer was an asset in such a course but that something was lost because the students were not able to participate in the preparation of the programs.

At the same time, a course on computing for social science students was introduced into the Hope curriculum. This course is basically a FORTRAN programming course. Those who were involved with teaching this course found that some knowledge of statistics would be valuable to those students enrolled. With some statistics background, the students could be assigned projects pertinent to their fields of interest.

It was therefore the consensus of the Hope College mathematics department that it would be more valuable to combine statistics and computer programming into one course for social science students rather than to continue with two separate courses. A proposal was made in the Fall of 1970 to the National Science Foundation for the development of such a course, in conjunction with the development of a laboratory for the mathematical probability and statistics course, and this project was funded.

Value of the computer. There are three major reasons that the computer is an asset to the introductory statistics course. First, exposure to the computer and computing is a necessary experience for any social science student. He should be aware of the economic, sociological and psychological impacts of the computer as well as the application of the computer to the solution of problems in his discipline.

Second, the computer can serve as an aid in teaching statistics theory. The student has a much greater mastery of a concept after he has explained it to someone else. When a student writes a program he is doing just this, explaining to the computer how to solve the problem. For example, in the assignment shown in Figure 3, the student is asked to explain to the computer how to test hypotheses. Also the computer can be used to provide the student with illustrations which further enhance his understanding. An example of this is found in the assignment given in Figure 2 in which the student is asked to illustrate the normal approximation to the binomial.

Third, the computer allows the student to apply the statistical procedures he is learning to useful sets of data, thus giving him valuable experience in interpreting results and an interesting incentive for learning the statistics.

<u>Description of the course</u>. The course being developed is entitled "Applied Statistics and Computer Programming." The only prerequisite is high school algebra. It is a two semster course for three hours credit each semester. An outline of the topics presented in the course is found in Figure 1.

The students are introduced to a simplified form of input and output so that they can begin programming without being exposed to FORMAT statements. This is done by subroutines written for this course because Hope College has an IBM 1130 which has no simplified input/output included in the system. The author has prepared notes to serve as a text for the class for the FORTRAN portion of the course because existing texts present the language in a sequence different from that determined to be optimal for this course. For example, we present subscripted variables very early in the course because they are needed to program examples and procedures in descriptive statistics.

The statistics text chosen is Elementary Statistics by Paul G. Hoel, and the topics covered follow the presentation of the text with exceptions noted below. Descriptive statistics and probability are reversed in order that the students may gain some familiarity with FORTRAN and subscripted variables before they are needed for descriptive statistics. Some discussion is added concerning random number generators along with experience in their use when random sampling is treated. Also, the students are given practice in using canned subroutines and interpreting their results for regression and analysis of variance.

Five data sets, each consisting of several variables, are stored on a disk. The students learn early in the course how to access this data. These sets are data actually used for research purposes in the areas of sociology, psychology, education and economics, and have been contributed by faculty on the Hope campus from their research and from other books and articles. Already three additional sets have been contributed for next year.

Each student is assigned one data set and one variable from that set which he uses throughout the year. This use ranges from finding the mean and standard deviation to taking random samples to obtain confidence intervals to correlation and regression with other variables in the same data set.

The students work fewer textbook problems than in the standard statistics course. Instead, they write computer programs for solving these problems and then apply these programs to their data sets. The students are given a total of 25 assignments involving the computer throughout the year. This averages out to slightly less than one assignment per week. The assignments typically involve the writing of a progam, the application of that program, and the answering of some questions intended to bring out the important points. In many cases students are asked to punch cards summarizing their results so that a mean result may be obtained for the entire class. Extra credit problems are given along with each assignment to challenge the better students. Two sample assignments are found in Figures 2 and 3.

Description of the project. The project for developing this course is of two years duration. In the summer of 1971 this course was organized with the assistance of four senior mathematics majors who wrote the necessary subroutines and programming examples as well as testing the computer assignments.

The course is being taught for the first time in the academic year 1971-72 with a starting enrollment of 26 students. The summer of 1972 will be devoted to revising the course according to the experience gained during the preceding academic year. The course will then be taught in revised form during the academic year 1972-73 and the following summer a final report will be made along with the final preparation of materials such as course outline, lecture notes and assignments. These materials will be distributed to allow other staff members to teach the course.

Results. At this writing it is the middle of the first year of the project and hence too early for any firm discussion of results. Thus far the reaction of the students has been most favorable. Some have indicated that they feel the statistics is made easier to understand by the use of the computer. I suspect, however, that there are others for whom the computer simply clouds the issue. More students have been completing the extra credit portion of the assignments than was expected.

The author feels that the morale and interest of the students is much greater in this course than in either of the two parent courses, and for this reason, the course is a pleasure to teach. There has also been a favorable response to this course from our social science faculty who feel it is a most valuable course and are encouraging their students to take it as well as assisting us in its development.

First Semester

1. Introduction to computers - Computers, algorithms, languages.
2. Elements of FORTRAN - Constants, variables, assignment statements, transfer statements, simplified input/output.
3. Probability - Rules of probability, counting techniquues
4. Subscripted variables
5. Descriptive statistics - Frequency distributions, mean, standard deviation, other measures.
6. Probability distributions - Discrete, continuous, Chebyshev's inequality, Binomial, Normal.
7. Random sampling - Random number generators, sample means.
8. Subprograms - Functions, subroutines.
9. Other computer languages and systems - COBOL, list processing languages and their applications, time-sharing, man-machine interaction.

Second Semester

1. Input/Output - FORMAT statements.
2. Estimation - Confidence intervals, t-distribution.
3. Testing Hypotheses - Type I and II errors, testing means, testing proportions.
4. Correlation and regression - Multiple linear regression, non-linear regression, interpretation of results.
5. Multiple subscripts
6. Chi-square - Contingency tables, goodness of fit.
7. Analysis of variance - use of library programs.
8. Nonparametric tests - Sign test, rank-sum test.
9. Miscellaneous - Implications of computers, experimental design, simulation.

Figure 1. Outline of the course.

Purpose: The purpose of this assignment is to illustrate how the normal distribution can be used as an approximation to the binomial, when the approximation is good, and how to use the normal table program XNORM.

Description: Write a program which reads n and p and using subprogram BINOM computes the binomial probability that $x = k$ for $k = 0,1,\ldots,n$, and the normal probability that $k - \frac{1}{2} \leq x \leq k + \frac{1}{2}$ for the normal distribution with mean np and variance $np(1-p)$, and the same values of k. Use XNORM described in Appendix E to find the normal probability.

Output: The output is to consist of one line containing n, one containing p, followed by n+1 lines each containing a value of k, the corresponding binomial probability and the normal probability.

Questions: 1. Try out your program for a variety of values of n and p. Do the cases where np and $n(1-p)$ are both greater than 5 show good accuracy? How is the accuracy when the above rule is violated?

2. Does the accuracy of the approximation tend to vary with k for fixed n and p? If so, how?

3. Do you notice that one probability is always larger than the other? Can you explain this?

Extra things to Try: Write a program which is the same as the one described above but which computes the probability that $x \leq k$ instead of $x = k$. Answer questions 1-3 for this program.

Figure 2. Sample Assignment

Normal Approximation to the Binomial

<u>Purpose</u>: The purpose of this assignment is to introduce the student to testing a hypothesis about a sample mean and illustrate type I and type II errors.

<u>Description</u>: Write a function subprogram which has the following arguments: XMU0, the hypothesized mean, XMU, the actual mean, SIG the actual standard deviation, and N, the sample size. The subprogram is to generate 100 samples of size N from a normal distribution with mean XMU and standard deviation SIG. For each sample, a 95% confidence interval is constructed assuming sigma known, and a test is made as to whether XMU0 is in the confidence interval, i.e., whether μ = XMU0 is accepted. A count is made of the number of times the hypothesis is accepted This is the value to be returned for the function.

Write a calling program which calls this function four times, each time with XMU0=20, SIG=5, N=10, and for XMU=20,22,25,30.

<u>Output</u>: Your output should consist of XMU0, XMU, SIG, N and the value of the function for each call of the function.

<u>Questions</u>: 1. Relate the results of each call to function to either type I or type II errors. Specify which.

2. Punch a card summarizing your results.

3. What would be the effect on your answers if SIG were 10 instead of 5? What if N were 20 instead of 10? What if we used a 99% instead of 95% confidence interval?

Figure 3. Sample Assignment

Testing of Hypotheses

4. Compute the theoretical probability of making an error in each of the four performed tests of hypotheses. Indicate for each whether it is a type I or a type II error. Compare these with your results.

5. The above program tests the hypothesis $\mu = 20$ against the two-sided alternative $\mu \neq 20$. How would your answers be changed if a one-sided alternative $\mu > 20$ were used? What if $\mu < 20$ were the alternative?

Extra things to try: Add enough generality to your program that you can try some of the things suggested in questions 3 and 5 above.

Figure 3 (cont.) Sample Assignment

FINAL TECHNICAL REPORT

Name of Institution:	Hope College
Address:	Holland, Michigan 49423
Principal Investigators:	Elliot A. Tanis Herbert L. Dershem
Grant Number:	GJ - 28786
Period Covered by Grant:	June, 1971 to August, 1973
Title:	Introduction of the Computer in the Statistics Curriculum

1) BRIEF DESCRIPTION OF PROJECT

The purpose of this grant was twofold: (i) Materials were developed for an integrated course in statistics and computer programming for social science students. The goal of the course is to use each of these two subjects as an aid in teaching the other. (ii) Materials were developed for a laboratory for a two semester junior-senior level mathematical probability and statistics course. In this laboratory the students have the opportunity to increase their understanding and appreciation of the theory by performing simulations on the computer.

During the summer of 1971 Professors Dershem and Tanis, along with four capable students, wrote exercises and developed supporting software for our IBM 1130 computer.

During the 1971-72 academic year the course changes were introduced into our curriculum. The four students continued to help us implement the changes.

Materials for both courses were revised during the summer of 1972. Three students assisted during this summer.

A conference was held on the Hope College campus August 14-18, 1972, which was attended by 52 college teachers of statistics from throughout the Midwest. At this conference the philosophy and the materials of the two courses developed under this grant were presented to the participants in formal sessions. During more informal sessions, the participants were invited to react to the materials and make suggestions for improvement. Copies of all materials developed, including exercises, program descriptions and program printouts, were given to all participants.

During the 1972-73 academic year, the two statistics courses were taught for the second time by Professors Dershem and Tanis using the modifications made during the preceding summer.

A final revision of the newly developed materials was made during the summer of 1973. Professor John Whittle assisted Professors Dershem and Tanis with this revision. The materials were collected into two separate manuals. One manual contains the computer assignments for the applied statistics and computer programming course. The other manual contains the exercises for the mathematical probability and statistics laboratory.

2) PUBLICATIONS

Tanis, Elliot A., "Theory of Probability and Statistics Illustrated by the Computer," Proceedings of The 1972 Conference on Computers in Undergraduate Curricula, Atlanta, Georgia, 1972, pp. 513-520.

Dershem, Herbert L., A Course on Computing and Statistics for Social Science Students," Proceedings of The 1972 Conference on Computers in Undergraduate Curricula, Atlanta, Georgia, 1972, pp. 525-528.

Tanis, Elliot A., "A Computer Laboratory for Mathematical Probability and Statistics," Proceedings of a Fourth Conference on Computers in the Undergraduate Curricula, Claremont, California, 1973, pp. 416-426.

Tanis, Elliot A. and Crisman, Roger, "Two Normal Distribution Problems with the Same Solution," submitted to The American Statistician.

Tanis, Elliot A., "Mathematical Probability and Statistics Computer Laboratory," submitted to the First British Conference on Computers in Higher Education.

3) REPORTS

Both Dr. Dershem and Dr. Tanis gave a report of this project at a Great Lakes Colleges Association Computing Symposium which was held March 7-8, 1972, at Wabash College.

Three of the papers listed above were presented at the Conferences on Computers in Undergraduate Curricula held in Atlanta in 1972 and in Claremont in 1973.

Dr. Dershem made a report of this project at the annual meeting of the Michigan Section of the Mathematical Association of America in May, 1973. At this same meeting Roger Crisman, one of our students, described his Senior Honors Project which made extensive use of the computer software developed in this project.

4) PERSONNEL

The following professors were associated with and supported by this grant:

Dr. Herbert L. Dershem, Associate Professor of Mathematics, Hope College.

Dr. Elliot A. Tanis, Professor of Mathematics, Hope College.

The following students received support during the summer of 1971 and the 1971-72 academic year:

Lynn Klaasen Hillegonds

Timothy Hillegonds

Richard Pohl

Glen Weener

The following students received support during the summer of 1972 and the 1972-73 academic year:

Roger Crisman

Richard Meyers

The following student received support during the summer of 1972:

Gail Ringsmith Buis

The following professor received support during the summer of 1973:

Professor John T. Whittle,

Assistant Professor of Mathematics, Hope College.

5) COMMENTS

Our computer center staff was very cooperative throughout this project. During the development of the computer software, the computer center contributed 190.471 hours of time at $30 per hour. This amounts to a contibution of $5,714.13.

As stated earlier, the 52 conference participants received a complete set of our materials. In addition we have sent materials to over 25 educators throughout this country and in two foreign countries.

During the current academic year two of our colleagues, Professors John Whittle and Jay Folkert, are using our materials. Their experiences will be helpful in future revisions.

Both Dr. Dershem and Dr. Tanis plan to submit reports to the National Technical Information Service of the Department of Commerce to facilitate wider dissemination of our results.

Hope College is currently installing a Xerox Sigma 6 computer which has time sharing capabilities. We plan to determine which of our materials are appropriate for use with terminals and then make the necessary adaptations.

We believe that both parts of this project have been successful. Dr. Dershem has applied to COMPUTe at Dartmouth for the purpose of writing up his laboratory materials for publication and hence wider dissemination. Dr. Tanis plans to publish a revision of his laboratory materials after the completion of a companion textbook which he is writing with Dr. Robert V. Hogg from the University of Iowa.

Elliot A. Tanis

Elliot A. Tanis

Herbert L. Dershem

Herbert L. Dershem

November 16, 1973

Date

FINAL BUDGET REPORT

Name of Institution:	Hope College Holland, Michigan 49423
Principal Investigators:	Elliot A. Tanis Herbert L. Dershem
Grant Number:	GJ-28786
Period covered by grant:	June 1971 to August 1973
Title:	Introduction of the Computer in the Statistics Curriculum

	Budget	Expenditures
A. SALARIES AND WAGES		
1. Principal Investigator - $\frac{1}{4}$ time one academic year; 5 months summer	$10,875	$11,135.00
2. Two Faculty Associates - $\frac{1}{4}$ time one academic year; 5 months summer and part-time one academic month	10,650	11,085.00
3. Pre-Baccalaureate Students (4)	6,120	5,757.36
4. Secretarial	500	0.0
TOTAL SALARIES AND WAGES	$28,145	$27,977.36
B. EXPENDABLE EQUIPMENT AND SUPPLIES	710	490.99
C. TRAVEL, Domestic - 50 conference participants @ avg. 500 miles/participant @ 8¢/mile	2,000	2,567.40
D. PER DIEM - 50 conference participants @ $10/day X 4 days	2,000	1,732.50
E. OTHER COSTS		
1. Publicity (conference) $ 200		
2. Reference Material 100		
3. Key Punch Rentals 3,120		
TOTAL OTHER COSTS	3,420	3,480.75
TOTAL DIRECT COSTS	$36,275	$36,249.00
F. INDIRECT COSTS - 45.3% of A. 1 and 2, as requested	9,751	9,751.00
TOTAL COSTS	$46,026	
AMOUNT OF AWARD - Rounded	$46,000	$46,000.00

Elliot A. Tanis

Elliot A. Tanis
Project Director

William K. Anderson

William K. Anderson
Controller and Chief Fiscal Officer