COMPUTER SIMULATION OF VARIANCE COMPONENTS ESTIMATES: PLOTS OF FREQUENCY DISTRIBUTIONS*

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

BU-245-M

June, 1967

Abstract

Frequency distributions of estimated between-group variance components derived from simulated data of a 1-way classification are compared with approximations to the exact density function of the estimators. Plots for a limited number of n-patterns are shown and discussed.

Details of the methods used are given in BU-233-M. Computing procedures are shown here.

---

* Hand-out for paper presented at 6th International Biometrics Conference, Sydney, Australia, August 1967.

COMPUTER SIMULATION OF VARIANCE COMPONENTS

ESTIMATES: PLOTS OF FREQUENCY DISTRIBUTIONS*

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

BU-245-M June, 1967

The six accompanying figures pertain to the distribution of between-group variance component estimates in a 1-way classification random model. Curves are shown in pairs: one is the frequency distribution of simulated values of each estimate, and the other is an approximation to the density function of that estimate, derived from Wang's (1966) distribution.

Description

The figures are for a limited number of n-patterns and values of the between-groups component σ_a^2, with σ_e^2 taken as unity at all times. Most cases are of just 5 groups (c = 5) and 25 observations (N = 25). One is for 45 observations in 5 groups and Figure VI is for 60 observations in 20 groups. Plots for more extensive n-patterns are in preparation.

All plots have been generated as computer output. For each there has been 2000 simulations of $\hat{\sigma}_a^2$, based on taking $\hat{\sigma}_e^2$ as a simulated χ_{N-c}^2-variable, and taking each $\bar{x}_i$ as $r_1\sigma_a + r_2/\sqrt{n_i}$ where r_1 and r_2 are pseudo-random variables distributed as N(0,1), simulated by a table look-up procedure from the medians of 500 equi-probable areas on each side of the mean of the standardized normal distribution (Searle, 1966). The approximation to the density function of $\hat{\sigma}_a^2$ has been derived by assuming that $\hat{\sigma}_a^2$ is a variable of the form $\alpha\chi_q^2 - \lambda_2\chi_{N-c}^2$ where λ_2 is known (Searle, 1966). In each case estimates of α

---

* Hand-out for paper presented at 6th International Biometrics Conference, Sydney, Australia, August 1967.

and q have been obtained by equating the first and second moments of $\hat{\sigma}^2_a$ and $\alpha\chi^2_q - \lambda_2\chi^2_{N-c}$, and with these values for α and q points on the density function of $\alpha\chi^2_q - \lambda_2\chi^2_{N-c}$ have been computed using Wang, (1966).

Headings on the graphs show: the n-pattern used (the values of the n_i's); the value of σ^2_a used in the simulation, denoted by A; the variance of σ^2_a Searle (1956), denoted by VAR(A); and the mean and variance of the 2000 values of $\hat{\sigma}^2_a$ obtained from the simulations. To facilitate computer generation abscissae are measured on the vertical rather than the horizontal axis. Fifty-one intervals of finite width (0.1)SE are used, with center points at σ^2_a - 2SE through to σ^2_a + 3SE, where SE is the standard error of $\hat{\sigma}^2_a$. Tail intervals are from $-\infty$ to σ^2_a - 2.05SE and from σ^2_a + 3.05 to $+\infty$. The left-hand column on each graph shows the values of the center points of the intervals, and in most cases there is a right-hand column showing the cumulative frequency of the simulated values. Simulated frequencies are plotted with the symbol X and approximate density functions with *, the latter being used whenever the two coincide.

Discussion

The attached figures are to be considered as preliminary results only. Discussion of them is therefore brief. Figures I - V pertain to situations of only 5 groups, for which an n-pattern such as (5,5,5,5,5) denotes 5 observations in each group.

Figure I. This is the balanced case, (5,5,5,5,5), with $\sigma^2_a = \frac{1}{4}$, a relatively small value. The two curves appear to be quite similar and not unlike a χ^2 curve.

Figure II. This is the same balanced case, (5,5,5,5,5), as in I but with a considerably larger value for σ^2_a, namely $\sigma^2_a = 20$. The curves are quite similar to those for $\sigma^2_a = \frac{1}{4}$ shown in Figure I except for being a little steeper on the left.

Figure III. Six graphs are shown for a moderately unbalanced situation, (1,1,7,8,8), over a range of values for σ_a^2, namely $\frac{1}{4}$,1,2,5,10 and 20. The curves are still somewhat like a χ^2, with increasing steepness on the left for the larger values of σ_a^2 (e.g. section 6 where σ_a^2 = 20, compared to section 2 where σ_a^2 = 1). There also appears to be a tendency for the curves of the simulated values to be slightly 'squeezed' compared to those of the approximate theoretical densities (sections 4 and 5, for example). By this we mean that the curves are somewhat as follows.

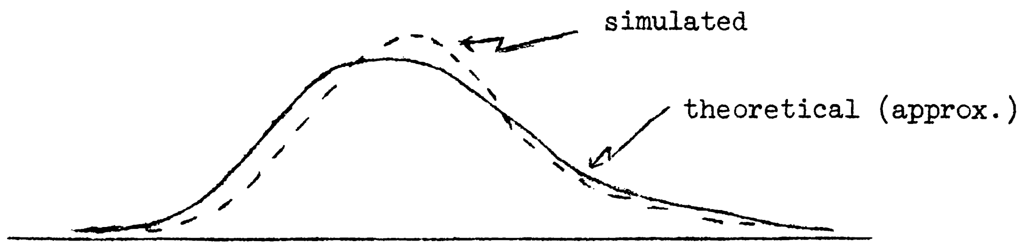


Figure IV. The four graphs here are for a very unbalanced case, (1,1,1,11,11), with $\sigma_a^2 = \frac{1}{4}$, 1, 5 and 20. The increasing steepness as σ_a^2 increases is very noticeable, the curves for σ_a^2 = 5 and 20 being almost exponential in type, corresponding to the approximate degrees of freedom, q, being close to unity, 1.91 and 1.83 respectively.

Figure V. Four cases representing increasing unbalancedness, all with σ_a^2 = 1: (5,5,5,5,5), the balanced case; (1,1,3,10,10), moderate unbalancedness; (1,1,1,1,21), very unbalanced, these three all having 5 groups and 25 observations; and in section 4, (1,1,1,21,21), representing both severe unbalancedness and the addition of 20 observations compared to (1,1,1,1,21) shown in 3. The trend for these cases seems clear: as unbalancedness increases there is increasing steepness on the left, with (1,1,1,21,21) being like an exponential, again corresponding to $q < 2$, in this case $q = 1.41$.

Figure VI. This shows the unbalanced case of (1,1,1,... for 19 groups, 41), for 60 observations in 20 groups, with four values of $\sigma_a^2 = \frac{1}{4}$, 1, 5 and 20. Two points of interest can be noticed: (i) the simulated curves are 'squeezed' considerably compared to the theoretical approximations, to the point of the latter appearing to be quite a poor fit; (ii) the values of σ_a^2 are the same as those used with (1,1,1,11,11) in Figure IV, where for $\sigma_a^2 = 20$ the curves are exponential in type: but they are not in Figure VI.

Conclusion

No definitive conclusions seem feasible from these preliminary results. Hopefully, extensions will yield more concrete evidence of the effect of unbalancedness on the distribution of the between-groups variance component estimate in a between and within groups analysis of variance.

Acknowledgements

Grateful thanks go to E. C. Townsend who did all the computer programming, and to Y. Y. Wang for assistance in adapting her density function.

Appendix: Computing techniques

Simulations. The simulation of a χ^2 variable for $\hat{\sigma}_e^2$ is discussed in Searle (1967); and the table look-up procedure on medians of 500 equi-probable areas on each side of the mean of the standardized normal distribution is detailed in Searle (1966).

Wang's density function

Wang (1966) considers the variable

$$Z = \alpha\chi^2_{2n} - \beta\chi^2_{2m} \qquad \text{--- (1)}$$

where α and β are constants and χ^2_{2n} and χ^2_{2m} are two independent χ^2-variables with 2n and 2m degrees of freedom, respectively, n and m being integers. The density function is defined in two parts, for Z negative and for Z positive. With the constant

$$K_2 = \frac{\alpha^{m-1}\,\beta^{n-1}}{2^{n+m}\,\Gamma(n)\,\Gamma(m)(\alpha+\beta)^{n+m-1}} \qquad \text{--- (2)}$$

the part of the function for negative Z is

$$f_-(z) = K_2 e^{\frac{z}{2\beta}} \int_0^\infty e^{-\frac{1}{2}t}\,t^{n-1}[t - z(\tfrac{1}{\alpha} + \tfrac{1}{\beta})]^{m-1}dt \qquad \text{--- (3)}$$

and the part for positive Z is

$$f_+(z) = K_2 e^{\frac{-z}{2\alpha}} \int_0^\infty e^{-\frac{1}{2}t}\,t^{m-1}[t + z(\tfrac{1}{\alpha} + \tfrac{1}{\beta})]^{n-1}dt\ . \qquad \text{--- (4)}$$

When m is integer, (3) can be expressed as a finite sum after binomial expansion of the term in z, and so can (4) when n is integer.

In applying (2) and (3) to the approximation

$$\hat{\sigma}_a^2 = \alpha\chi^2_q - \lambda_2\chi^2_{N-c}\ , \qquad \text{--- (5)}$$

where λ_2 is known [equation (5) of Searle (1967)] we have, analagous to (1), $n = \frac{1}{2}q$ and $m = \frac{1}{2}(N-c)$. Now $\frac{1}{2}(N-c)$ is half the difference between the number of observations and number of groups and this can always, with little loss of generality, be made an even integer in planning simulation studies. This enables (3) to be simplified by binomial expansion, as just indicated. However, in using (5), α and q are given values derived from equating the first and second moments of both sides thereof, and so $n = \frac{1}{2}q$ is seldom an even integer. Hence (4) must be evaluated other than by binomial expansion. Details are shown below.

First the part for negative Z: from Wang (1966) the expansion of (3) is

$$f_-(z) = \left(\frac{\alpha}{\alpha+\beta}\right)^{m-1}\left(\frac{\beta}{\alpha+\beta}\right)^{n-1}\frac{1}{2(\alpha+\beta)}\, e^{\frac{z}{2\beta}} \sum_{j=0}^{m-1}\left(\frac{-z}{2\beta}\right)^{j}\left(\frac{\alpha+\beta}{\alpha}\right)^{j}\frac{(m+n-2-j)!}{j!(m-1-j)!(n-1)!}$$

This is a finite sum because

$$m = \tfrac{1}{2}(N - c) \text{ is an integer}$$

although

$$n = \tfrac{1}{2}q \text{ is not an integer.}$$

Now write

$$K = \left(\frac{\alpha}{\alpha+\beta}\right)^{m-1}\left(\frac{\beta}{\alpha+\beta}\right)^{n-1}\frac{1}{2(\alpha+\beta)} \equiv 2^{m+n-1}\Gamma(n)\Gamma(m)K_2$$

and $$u_j = \left[\frac{-z(\alpha+\beta)}{2\alpha\beta}\right]^{j}\frac{(m+n-2-j)!}{j!(m-1-j)!(n-1)!}\;;$$

even though n is not integer u_j has a finite number of terms because, with n being integer

$$\frac{(m+n-2-j)!}{(n-1)!} = (m+n-2-j)(m+n-3-j)\ldots(n+1)n\;.$$

Thus

$$f_-(z) = Ke^{\frac{z}{2\beta}}\sum_{j=0}^{m-1}u_j\;.$$

For computing purposes recurrence relationships among the u_j can be used:

$$u_0 = \frac{(m \quad n - 2)!}{(m - 1)!(n - 1)!} = \frac{(m + n - 2)(m + n - 3) \ldots (n + 1)n}{(m - 1)!}$$

and
$$u_j = \left[\frac{-z(\alpha + \beta)}{2\alpha\beta}\right] \frac{(m - j)}{j(m + n - 1 - j)} u_{j-1} \text{ for } j \geq 1 .$$

In passing, note from (3) that for $z = 0$

$$f(0) = K_2 e^{\frac{z}{2\beta}} \int_0^\infty e^{-\frac{1}{2}t} t^{m+n-2} dt$$

$$= K_2 \Gamma(m + n - 1) 2^{m+n-1}$$

$$= \frac{K\Gamma(m + n - 1)}{\Gamma(n)\Gamma(m)}$$

$$= Ku_0 .$$

Therefore on writing

$$v_0 = f(0) = Ku_0 = \frac{K(m + n - 2)(m + n - 3) \ldots (n + 1)n}{(m - 1)(m - 2) \ldots . 2. 1.}$$

and
$$v_j = \left[\frac{-z(\alpha + \beta)}{2\alpha\beta}\right] \frac{(m - j)}{j(m + n - 1 - j)} v_{j-1} \text{ for } j \geq 1$$

$$f_-(z) = e^{\frac{z}{2\beta}} \sum_{j=0}^{m-1} v_j .$$

The other part of the density function is for positive Z: here we use (4), but cannot invoke binomial expansion because n is not integer and t takes all values from 0 to ∞. Instead we write

$$r = \tfrac{1}{2}z\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) \qquad \text{--- (6)}$$

and use the transformation

$$t = 2r\tan^2\theta .$$

Then (4) becomes

$$f_+(z) = K_2 e^{\frac{-z}{2\alpha}} \int_0^{\frac{\pi}{2}} e^{-r\tan^2\theta}(2r\tan^2\theta)^{m-1}(2r\tan^2\theta + 2r)^{n-1}2r\,d(\tan^2\theta)$$

$$= K_2 e^{\frac{-z}{2\alpha}} \int_0^{\frac{\pi}{2}} e^{-r\tan^2\theta}(2r)^{m+n-1}(\tan^2\theta)^{m-1}(\sec^2\theta)^{n-1}2\tan\theta\sec^2\theta\,d\theta$$

$$= 2K_2(2r)^{m+n-1}e^{\frac{-z}{2\alpha}} \int_0^{\frac{\pi}{2}} e^{-r\tan^2\theta}(\tan\theta)^{2m-1}(\sec\theta)^{2n}d\theta \ .$$

Now on substitution from (2) and (6)

$$2K_2(2r)^{m+n-1} = 2K_2\left[\frac{2z(\alpha + \beta)}{2\alpha\beta}\right]^{m+n-1}$$

$$= (\tfrac{1}{2}z)^{m+n-1}\frac{\alpha^{m-1}\beta^{n-1}}{(\alpha\beta)^{m+n-1}\Gamma(n)\Gamma(m)}$$

$$= \frac{(\frac{1}{2}z)^{m+n-1}}{\alpha^n\beta^m\Gamma(n)\Gamma(m)}$$

and so $f_+(z)$ is

$$f_+(z) = \frac{(\frac{1}{2}z)^{m+n-1}e^{\frac{-z}{2\alpha}}}{\alpha^n\beta^m\Gamma(n)\Gamma(m)}\int_0^{\infty}\exp\left[\frac{-z(\alpha + \beta)\tan^2\theta}{2\alpha\beta}\right]\left(\frac{\sin^{2m-1}\theta}{\cos^{2m+2n-1}\theta}\right)d\theta \ .$$

In this form was $f_+(z)$ computed, using quadrature for the infinite integral, the simple trapezoidal method with 200 intervals between 0 and $\pi/2$.

In all cases the values of $f(z)$ computed by either of these means were multiplied by L/50, L being the length of the range in which the distribution is represented on the graph, name 5SE, where SE is the standard error of $\hat{\sigma}_a^2$. Thus the multiplier is (0.1)SE.

References

Searle, S. R. (1956). Matrix methods in variance and covariance components analysis. Ann. Math. Stat. 27, 737.

Searle, S. R. (1966). Properties of certain discrete distributions for generating approximately normal variables. Paper No. BU-228-M, mimeo series, Biometrics Unit, Cornell University, Ithaca, N.Y.

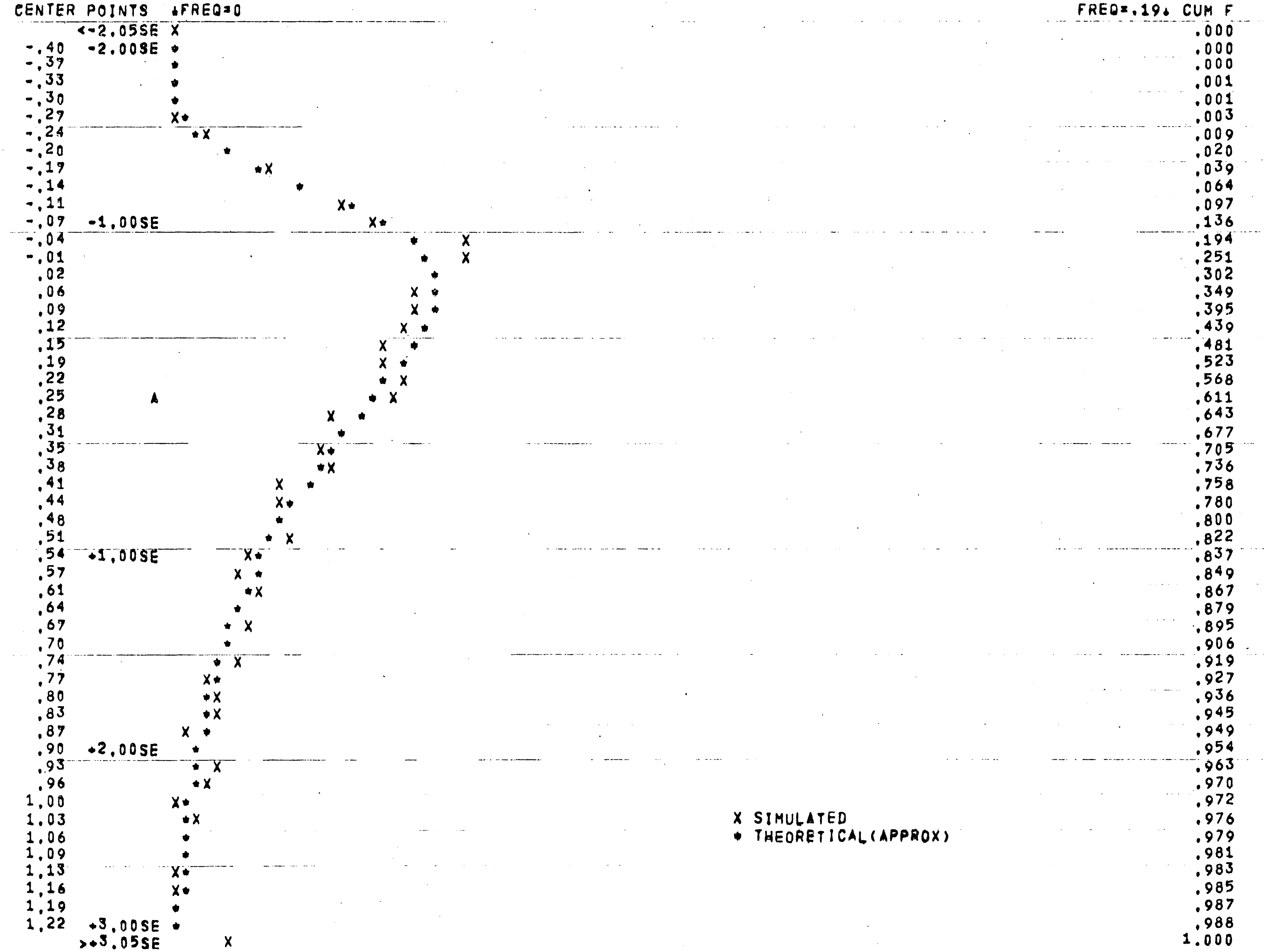
Searle, S. R. (1967). Computer simulation of variance components estimates. Conference Pre-pring, VIth International Biometric Conference, Sydney, Australia, August, 1967 (Paper BU-233-M in the mimeo series, Biometrics Unit, Cornell University.)

Wang, Y. Y. (1966). The distributions and moments of some variance component estimators. Biometrika (in press) .

FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANCE COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE

N-PATTERN; 5 5 5 5 5

POP. PARAMETERS;A= .25,VAR(A)= .11. 2000 SIMULATIONS YIELDED MEAN(A)= .251,VAR(A)= .106.



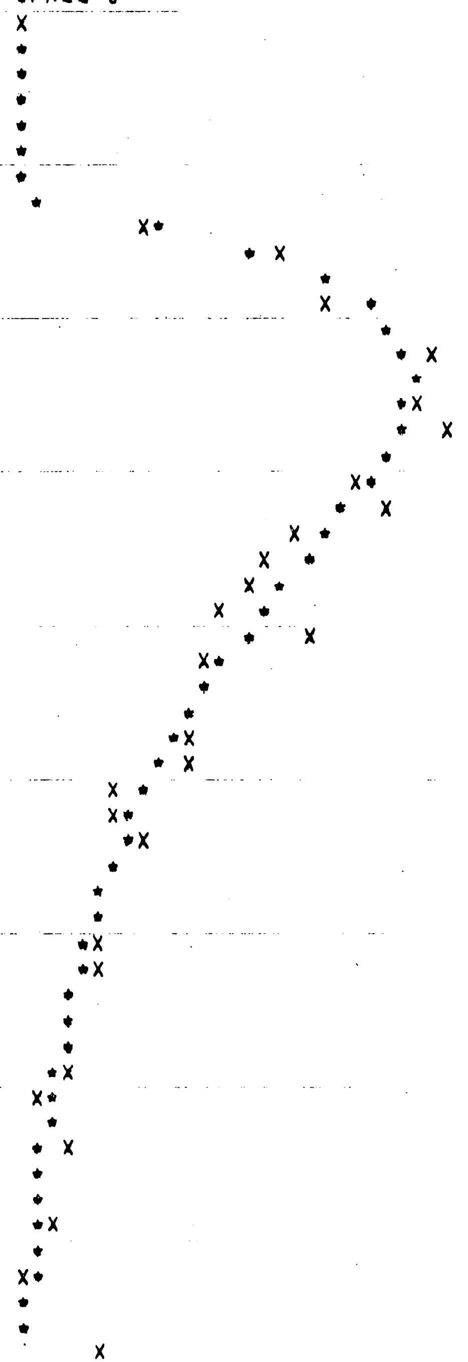
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANCE COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE

II

N-PATTERN; 5 5 5 5 5

POP. PARAMETERS;A= 20.00,VAR(A)= 204.02, 2000 SIMULATIONS YIELDED MEAN(A)= 19.992,VAR(A)= 196.952.

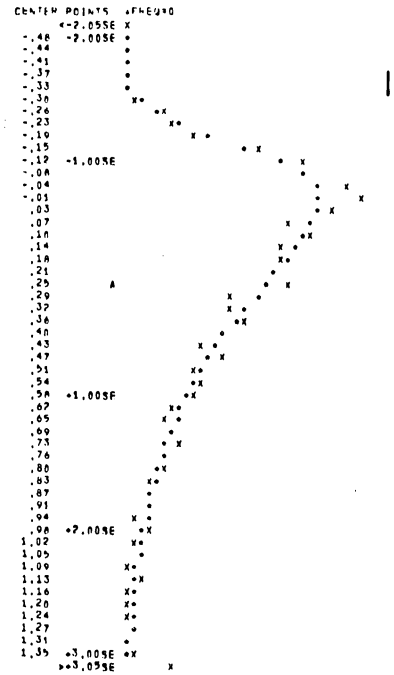
CENTER POINTS		+FREQ=0	FREQ=.19+ CUM F
	<-2.05SE		0
-8.57	-2.00SE		0
-7.14			0
-5.71			0
-4.28			0
-2.85			0
-1.43			0
.00			.003
1.43			.019
2.86			.053
4.29			.093
5.72	-1.00SE		.134
7.14			.183
8.57			.239
10.00			.292
11.43			.344
12.86			.401
14.29			.450
15.71			.497
17.14			.546
18.57			.583
20.00		A	.616
21.43			.646
22.86			.674
24.29			.713
25.71			.737
27.14			.762
28.57			.784
30.00			.806
31.43			.829
32.86	+1.00SE		.842
34.28			.855
35.71			.872
37.14			.884
38.57			.894
40.00			.904
41.43			.916
42.85			.927
44.28			.934
45.71			.941
47.14			.948
48.57	+2.00SE		.955
50.00			.958
51.42			.963
52.85			.970
54.28			.973
55.71			.976
57.14			.981
58.57			.985
59.99			.986
61.42			.987
62.85	+3.00SE		.989
	>+3.05SE		1.000



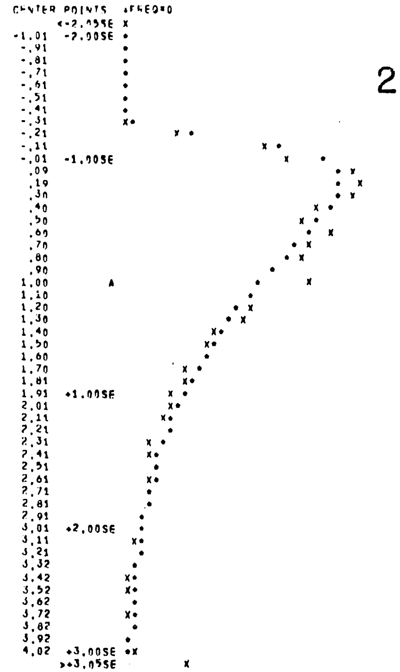
X SIMULATED
* THEORETICAL(APPROX)

III

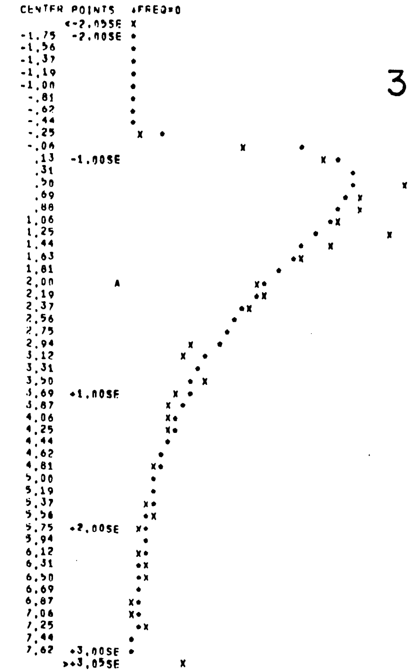
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC
N-PATTERN) 1 1 7 8 9
POP. PARAMETERS(A= .25,VAR(A)= .13. 2000 SIMULATIO S



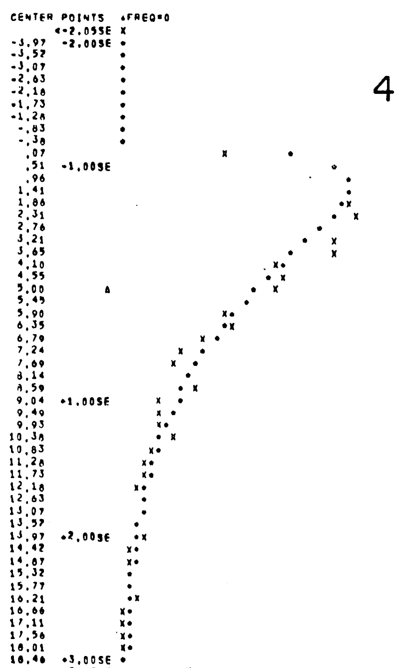
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC
N-PATTERN) 1 1 7 8 8
POP. PARAMETERS(A= 1.00,VAR(A)= 1.01. 2000 SIMULATIO S



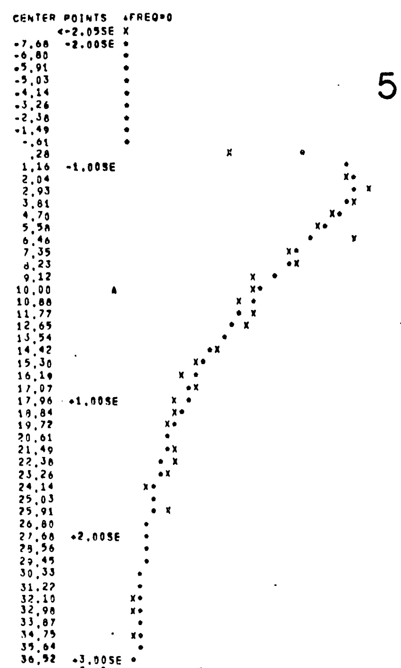
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC-
N-PATTERN) 1 1 7 8 8
POP. PARAMETERS(A= 2.00,VAR(A)= 3.91. 2000 SIMULATIO S



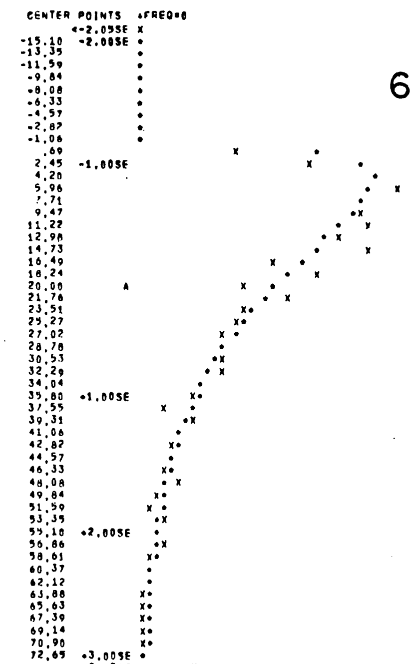
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC-
N-PATTERN) 1 1 7 8 8
POP. PARAMETERS(A= 5.00,VAR(A)= 20.12. 2000 SIMULATIO S



FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC- I
N-PATTERN) 1 1 7 8 8
POP. PARAMETERS(A= 10.00,VAR(A)= 78.15. 2000 SIMULATIO S

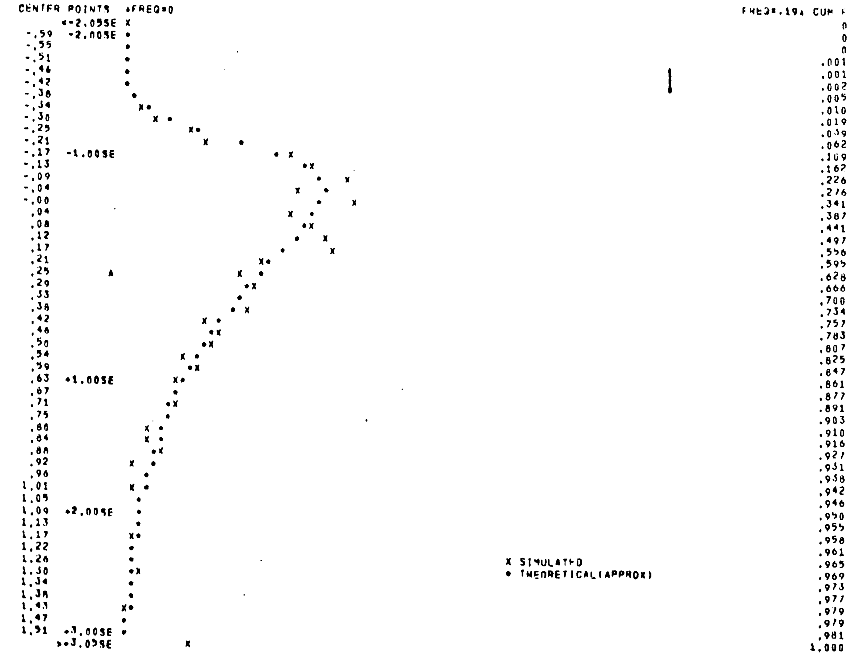


FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC
N-PATTERN) 1 1 7 8 8
POP. PARAMETERS(A= 20.00,VAR(A)= 388.02. 2000 SIMULATIO S

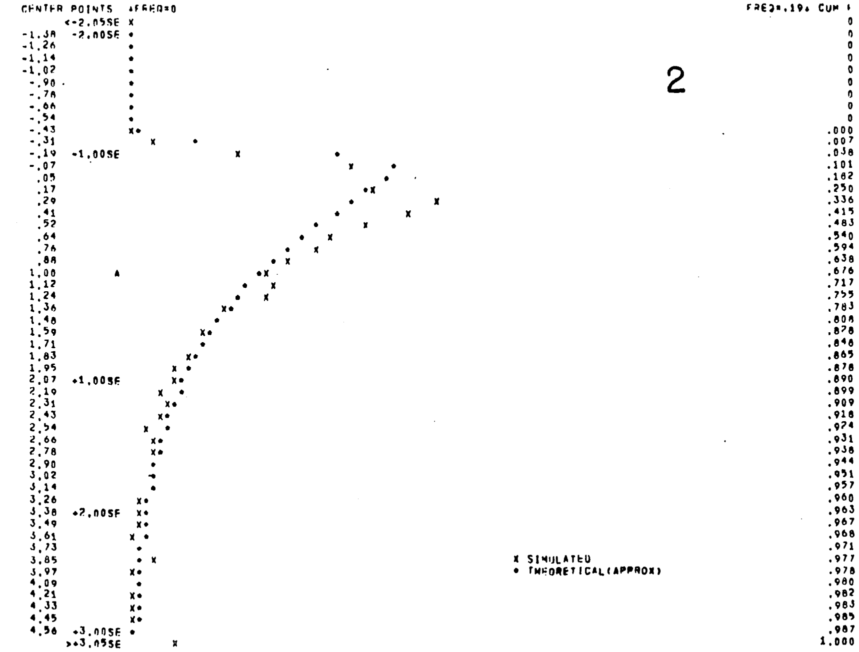


IV

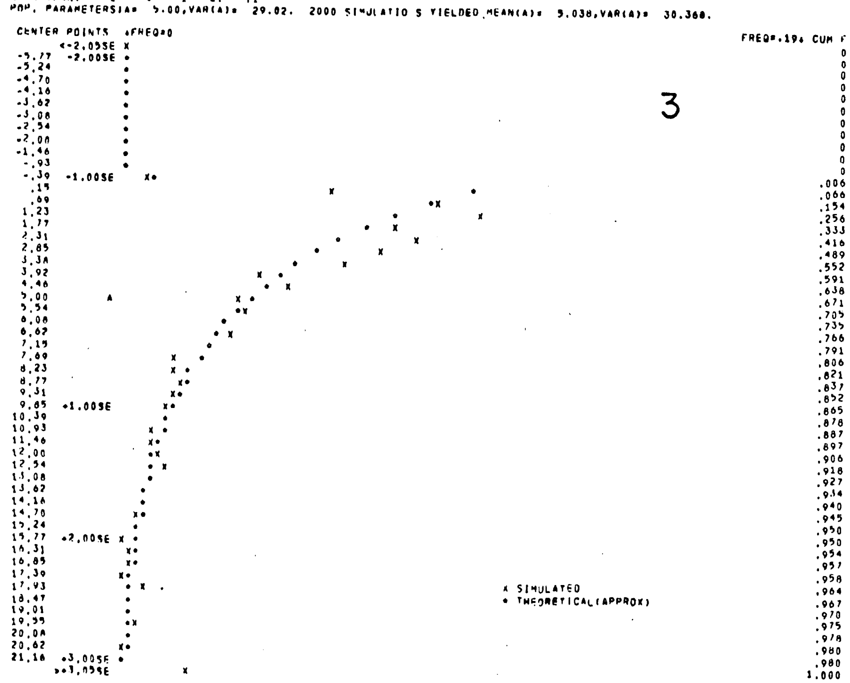
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 11 11
POP. PARAMETERS(A= .25,VAR(A)= .18. 2000 SIMULATIO S YIELDED MEAN(A)= .252,VAR(A)= .196.



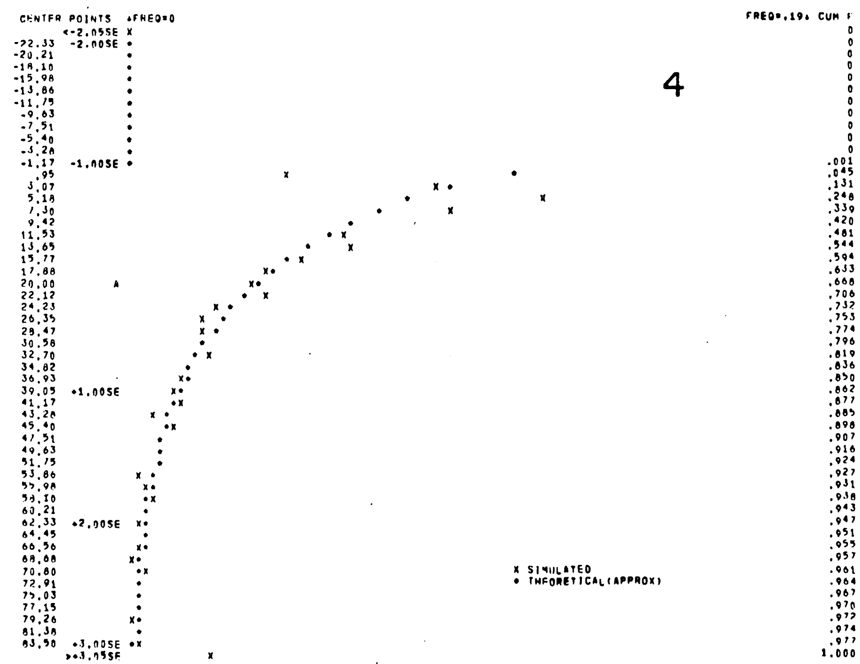
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 11 11
POP. PARAMETERS(A= 1.00,VAR(A)= 1.41. 2000 SIMULATIO S YIELDED MEAN(A)= .928,VAR(A)= 1.163.



FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 11 11
POP. PARAMETERS(A= 5.00,VAR(A)= 29.02. 2000 SIMULATIO S YIELDED MEAN(A)= 5.038,VAR(A)= 30.360.

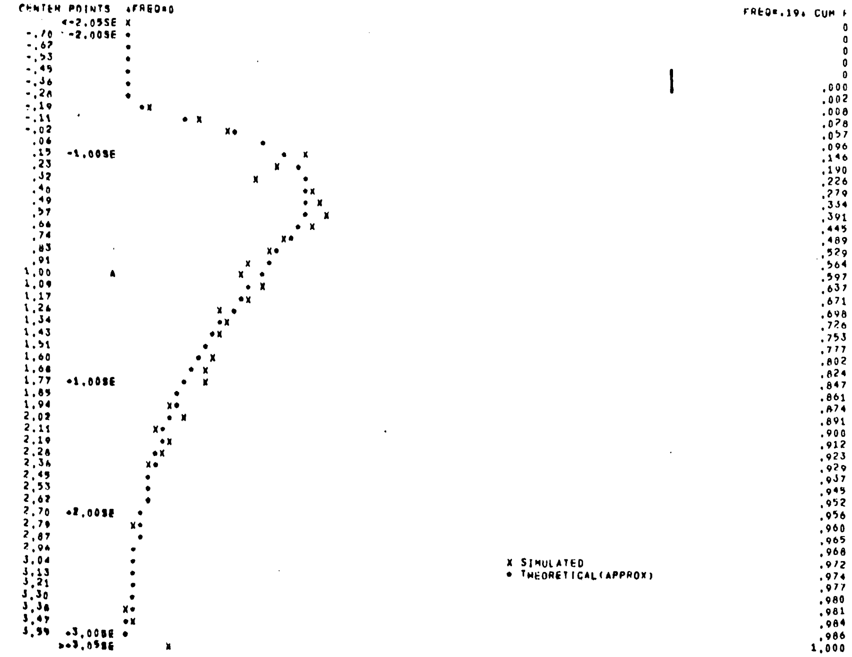


FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 11 11
POP. PARAMETERS(A= 20.00,VAR(A)= 447.96. 2000 SIMULATIO S YIELDED MEAN(A)= 20.604,VAR(A)= 478.604.

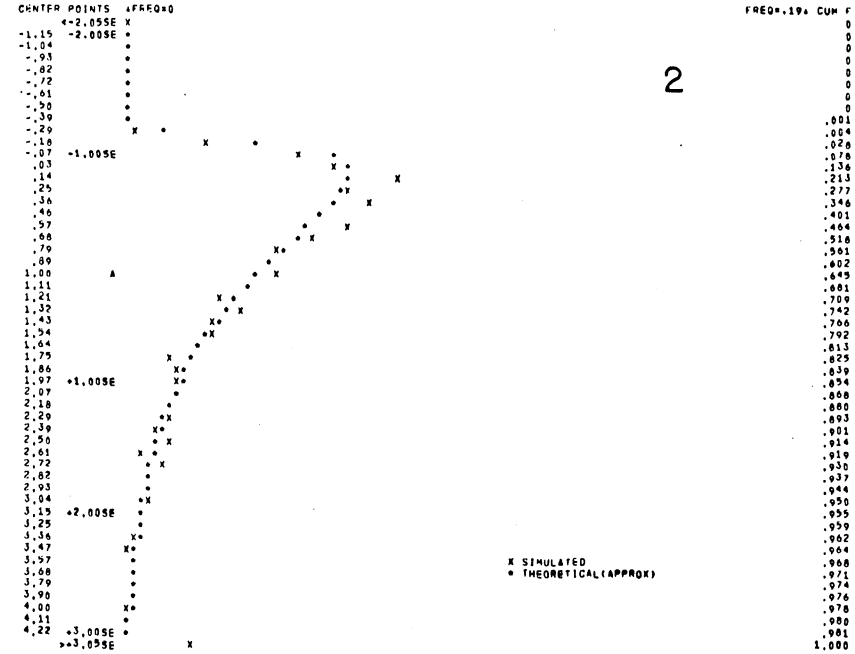


V

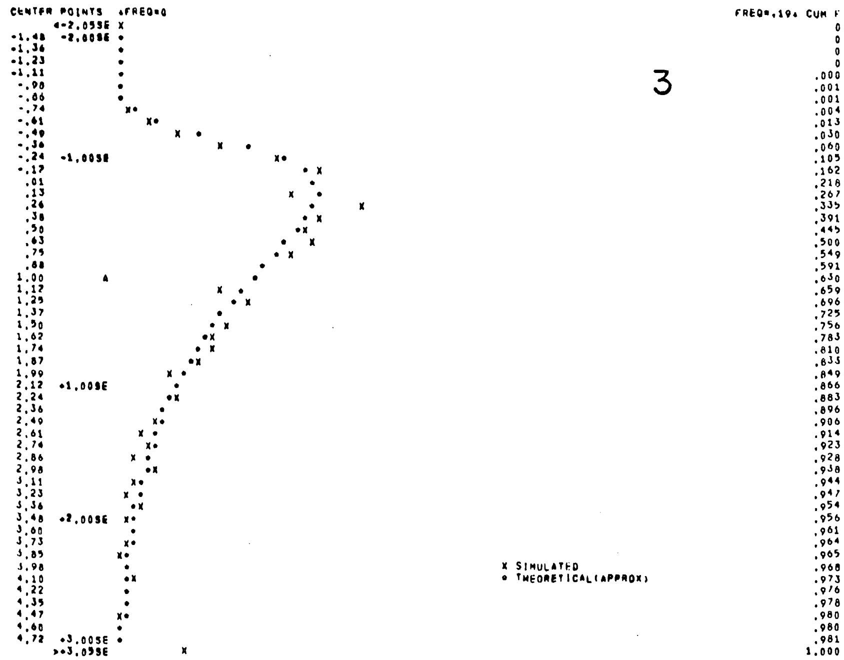
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC- COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 5 5 5 5 5
POP. PARAMETERS(A= 1.00,VAR(A)= .72. 2000 SIMULATIO S YIELDED MEAN(A)= 1.012,VAR(A)= .730.



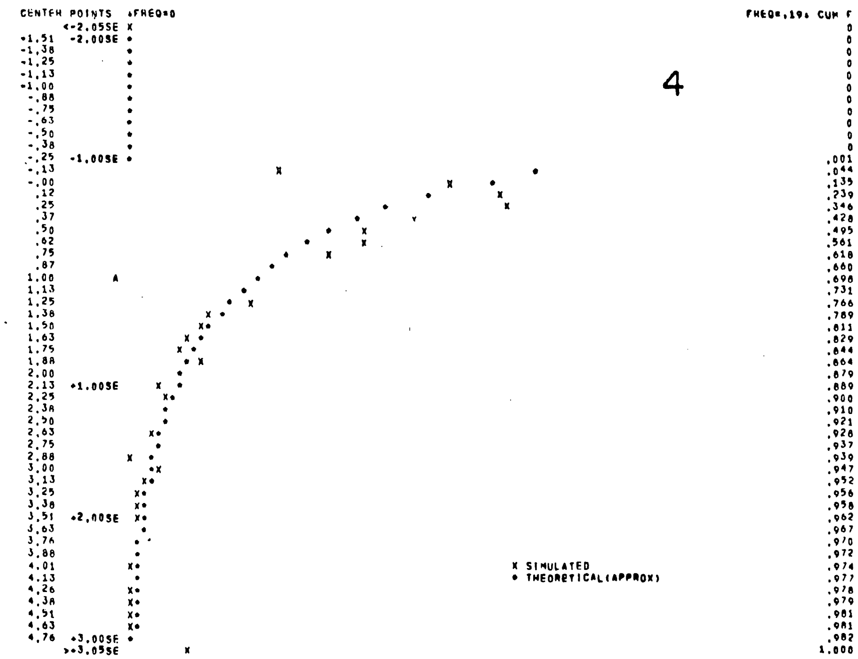
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 3 10 10
POP. PARAMETERS(A= 1.00,VAR(A)= 1.15. 2000 SIMULATIO S YIELDED MEAN(A)= 1.009,VAR(A)= 1.171.



FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 1 21
POP. PARAMETERS(A= 1.00,VAR(A)= 1.54. 2000 SIMULATIO S YIELDED MEAN(A)= .995,VAR(A)= 1.531.



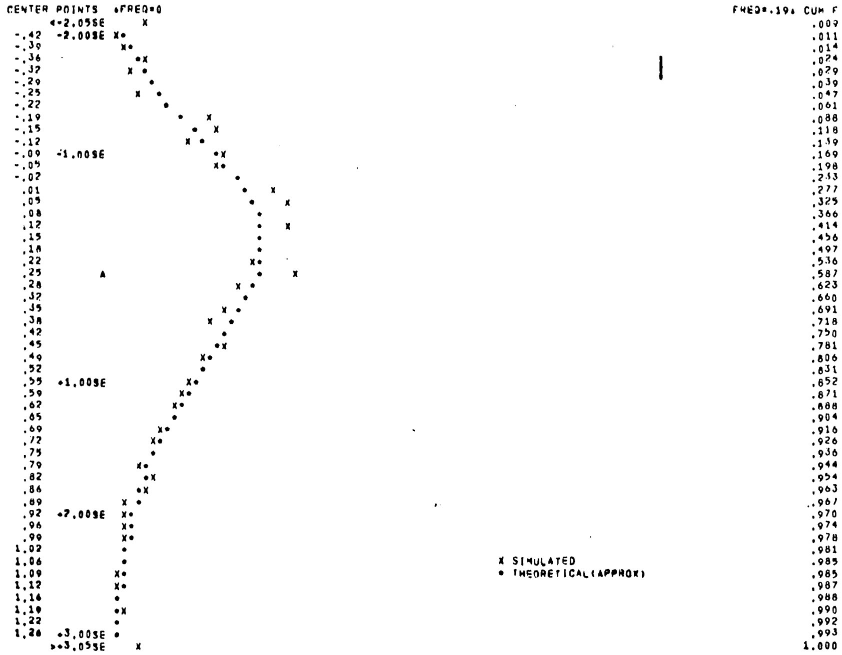
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN) 1 1 1 21 21
POP. PARAMETERS(A= 1.00,VAR(A)= 1.57. 2000 SIMULATIO S YIELDED MEAN(A)= .935,VAR(A)= 1.336.



VI

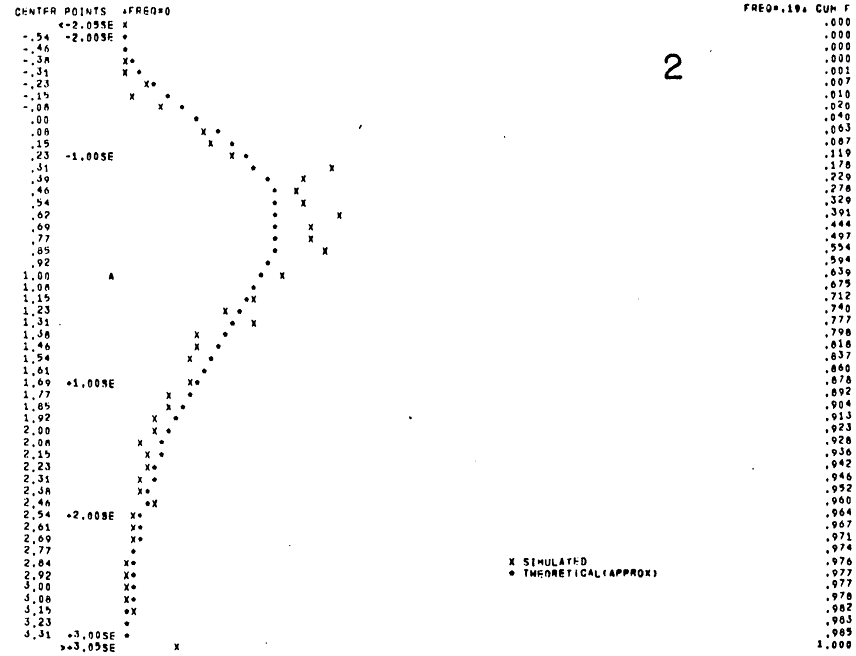
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 41

POP. PARAMETERS: .25, VAR(A)= .11. 2000 SIMULATIO S YIELDED MEAN(A)= .235, VAR(A)= .109.



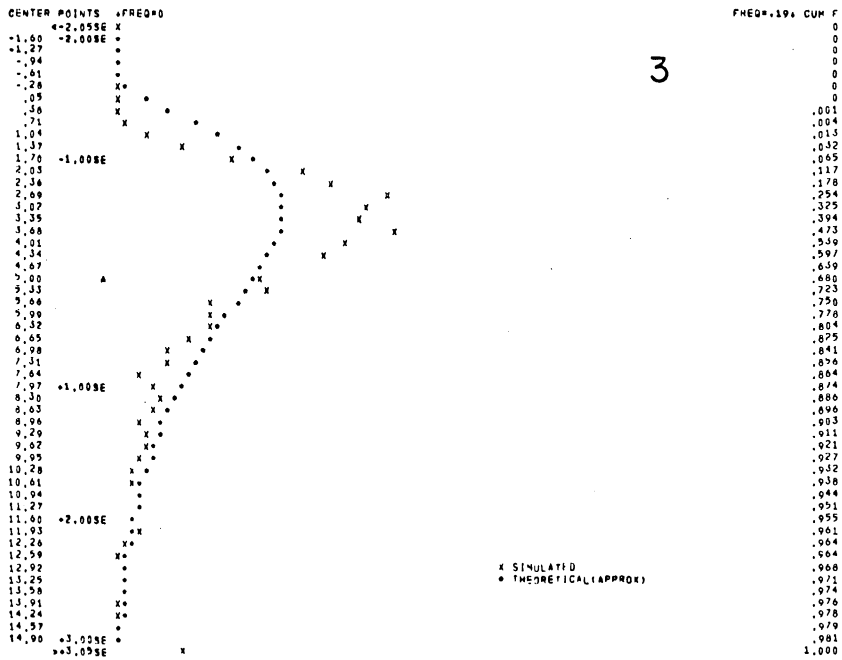
FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 41

POP. PARAMETERS: 1.00, VAR(A)= .59. 2000 SIMULATIO S YIELDED MEAN(A)= .962, VAR(A)= .552.



FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC- COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 41

POP. PARAMETERS: 5.00, VAR(A)= 10.90. 2000 SIMULATIO S YIELDED MEAN(A)= 4.907, VAR(A)= 10.931.



FREQUENCY DISTRIBUTION OF THE AMONG GROUPS VARIANC COMPONENT ESTIMATED FROM A 1-WAY ANALYSIS OF VARIANCE
N-PATTERN: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 41

POP. PARAMETERS: 20.00, VAR(A)= 165.95. 2000 SIMULATIONS YIELDED MEAN(A)= 20.154, VAR(A)= 163.377.