NASA Contractor Report 3633

A Study of the Feasibility of Statistical Analysis of Airport Performance Simulation

Raymond H. Myers

NOVEMBER 1982

NASA

NASA Contractor Report 3633

A Study of the Feasibility of Statistical Analysis of Airport Performance Simulation

Raymond H. Myers
Raymond H. Myers
Blacksburg, Virginia

Prepared for
Langley Research Center

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1982

A Study of the Feasibility of Statistical Analysis of Airport Performance Simulation

I. Introduction

The main objectives of studies planned by FAA/NASA implicitly require an adequate definition of airport capacity with view toward sorting out the roles of pilot, ground controller, and other factors that might influence the overall safety of our aviation system. Any future proposed simulation studies by NASA and FAA should, of course, be carried out with the aid of sound statistical tools. The study described here is a fundamental one in which the feasibility of conducting statistical analysis of such simulation experiments is investigated.

At the outset of the study, we attempt to study the form of the distribution of airport capacity in the standard simulation model (Ref. (2)). This capacity measure would be the basic response in any planned simulation experiment designed to study the roles of such variables as number of pilots, number of controllers, number of planes, environmental conditions, etc. on airport performance.

After the distribution of throughput capacity is assessed, an attempt is made to determine how effective standard experimental design and analysis of variance techniques would be in detecting capacity changes when conditions vary. In particular, the crucial problem here is to first determine the effect of the non-Gaussian Distribution of capacity on standard analysis of variance techniques and power calculations. Adjustments must be made in power calculations to account for the natural heterogeneity of variance in capacity. Secondly, power computations must be made in order to determine how economic simulation experiments would be if they are designed to detect capacity changes from condition to condition. In the latter, we are preoccupied with determining the number of replications in the simulation experiments that will result in an effective study, i.e., one that is able to detect small or moderate changes in mean capacity.

Many of the conclusions drawn here are a result of statistical simulation studies, i.e., studies in which significance level and power of standard tests are empirically studied on the computer via Monte-Carlo Techniques, and non standard conditions (non Gaussian and hetereogeneous variance) are built into the simulation.

II. Capacity Measure

The capacity measure to be used is given by the airport throughput rate

$$C = \text{no. of vehicles/hour} \,, \tag{2.1}$$

Hereafter the measure C will denote the mean throughput, a population parameter often called capacity. The inter-arrival time, IAT, represents the time between touchdowns. Because of the procedure inherent to the simulation that "buffers" the touchdowns, one can realistically assume that IAT follows a normal distribution with parameters (μ,σ). In addition, the buffer provides that μ and σ be related in the simulation through the following relationship (ref. 2)

$$\mu_{IAT} = \frac{3600s}{v} + Z_B \sigma_{IAT}$$

where

s = separation in nautical miles

v = approach velocity, knots

Z_B = buffer factor (Gaussian probability factor)

If $s = 2$, $v = 140$ and $Z_B = 1.96$, the following approximate relationships hold:

$$\mu_{IAT} = 51.4 + 1.96\sigma_{IAT} \quad \text{and}$$
$$C = \frac{3600}{\mu_{IAT}} \quad . \tag{2.2}$$

μ_{IAT} is measured in seconds. The capacities of interest in this study range from 20 to 60 vehicles/hour.

(a) Distribution of Capacity

It is of interest to initially consider the distribution of throughput rate, which of course is a random variable. Let IAT be a random variable denoted by y. Strictly speaking, y has a practical lower bound and thus follows a <u>truncated</u> <u>normal</u> <u>distribution</u>. For now, assume the point of truncation to be δ; thus the density of y is known to be

$$f(y) = K^{-1}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \qquad y \geq \delta \tag{2.3}$$

where

$$K = \int_{\delta}^{\infty} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

to begin, we need the distribution of the throughput $z = 1/y$. Following standard procedures

$$f(z) = K^{-1} e^{-1/2\sigma^2(\frac{1}{z}-\mu)^2} \left|\frac{\partial y}{\partial z}\right|$$

$$= K^{-1} e^{-1/2\sigma^2(\frac{1}{z}-\mu)^2} \left(\frac{1}{z^2}\right) \quad 0<z\leq\frac{1}{\delta} \qquad (2.4)$$

One can easily see that the above density, that of the inverse of a truncated normal, integrates to unity. For, we can write

$$\int_0^{1/\delta} f(z)dz = K^{-1} \int_0^{1/\delta} e^{-1/2\sigma^2(\frac{1}{z}-\mu)^2} \frac{1}{z^2}\, dz \qquad (2.5)$$

and invoke the transformation $\frac{1}{z} = w$, allowing $dz = -z^2 dw$. Thus (2.5) is written

$$\int_\infty^\delta K^{-1}\, e^{-\frac{1}{2}(\frac{w-\mu}{\sigma})^2}\, w^2(-\frac{1}{w^2})dw \quad .$$

But $\int_\infty^\delta e^{-\frac{1}{2}(\frac{w-\mu}{\sigma})^2} dw = -K$ and thus (2.5) equals unity.

Figures I - VI give an indication of the appearance of the distribution of z. Note of course that in many cases the throughput is far from Gaussian.

It is of interest to work out the moments of the distribution of z and also attempt to gain some insight on the distribution shape. An approximation for the variance of the distribution can be helpful in what follows. For the first moment

$$E(z) = K^{-1} \int_0^{1/\delta} \frac{1}{z}\, e^{-1/2\sigma^2(\frac{1}{z}-\mu)^2} dz \ .$$

Allowing $\frac{1}{z} = v$, we have

$$E(z) = K^{-1} \int_{\infty}^{\delta} v[e^{-1/2\sigma^2(v-\mu)^2}] - \frac{1}{v^2}\, dv$$

$$= K^{-1} \int_{\delta}^{\infty} \frac{1}{v}\, e^{-1/2\sigma^2(v-\mu)^2}\, dv$$

Similarly, one can show that

$$E(z^2) = K^{-1} \int_{\delta}^{\infty} \frac{1}{v^2}\, e^{-1/2\sigma^2(v-\mu)^2}\, dv$$

The above suggest that in certain situations, a simple approximation of $Var(z) = \sigma_z^2$ might be found in terms of the parameters of inter-arrival time, namely μ and σ. A Taylor Series approximation approach should be quite good for cases where $\mu >> \sigma$. For example, expanding $1/y^2$ and $1/y$ in a Taylor Series around $y = \mu$,

$$\frac{1}{y^2} \cong \frac{1}{\mu^2} - 2\frac{(y-\mu)}{\mu^3} + \frac{6(y-\mu)^2}{2!\ \mu^4} + \ldots$$

$$\frac{1}{y} \cong \frac{1}{\mu} - \frac{(y-\mu)}{\mu^2} + \frac{2(y-\mu)^2}{2!\ \mu^3} + \ldots$$

Thus we can approximate σ_z^2 by $E(\frac{1}{y})^2 - E[(\frac{1}{y})]^2$ and μ_z which gives

$$\mu_z \cong \frac{1}{\mu} + \frac{\sigma^2}{\mu^3}$$

$$\sigma_z^2 \cong \frac{1}{\mu^2} + \frac{3\sigma^2}{\mu^4} - (\frac{1}{\mu^2} + \frac{2\sigma^2}{\mu^4}) = \frac{\sigma^2}{\mu^4} \quad .$$

This type of approximation of inverse moments is not uncommon for cases where $\mu > 3\sigma$ (See Kendall & Stuart [1]). In our situation, our mean capacity range of 20 - 60 results in the following approximate ranges on μ and σ.

C(veh/hr)	μ_{IAT}(sec)	σ_{IAT}(sec)
60	60	5
51	70	10
45	80	15
40	90	20
36	100	25
30	120	35
25.7	140	45
20	180	65

It is clear that this approximation for Var(z) should be quite good until we reach the very low end of the capacity specifically when capacity is near 20 veh./hr. The approximation itself can be "checked" by simulation results. The purpose of the approximation lies in its use in developing an algorithm for computing power. This will be discussed in a later section.

III. Review of Analysis of Variance Principles

As we mentioned earlier, it is necessary to investigate a wide variety of conditions on mean throughput to determine if indeed one can effectively and economically compare capacity statistically from condition to condition with simulation experiments. The analysis of variance procedure results in obvious problems. These problems will be more obvious after we review analysis of variance procedures and the calculation of power of the analysis.

(a) Model and Assumptions

Briefly, when k group means are to be compared in an analysis of variance, a model is assumed of the type

$$z_{ij} = \gamma_i + \varepsilon_{ij} \qquad \begin{array}{l} i=1,2,\ldots,k \\ j=1,2,\ldots,n_i \end{array} \tag{3.1}$$

where γ_i is the group population mean, ε_{ij} is the disturbance or random error term of the model. The z_{ij} is the basic response measurement under the j^{th} observation in the i^{th} group. For our case, of course, z_{ij} would represent the j^{th} simulated throughput in the i^{th} group and γ_i the mean throughput or capacity for the i^{th}. One is interested in testing then

$$H_0: \ \gamma_1 = \gamma_2 = \cdots = \gamma_k$$

H_1: There is a difference in the means

The analysis of variance procedure then (see Walpole and Myers [2]) provides a methodology for testing this hypothesis. The usual test statistic is an F-statistic. One can consult [2] for the details. Actually, in our application it is quite likely that the model in equation (3.1) could and should be extended to cover the study of several groups simultaneously. For example in a so-called two way classification, one might want to simultaneously study number of pilots and number of ground controllers (and perhaps interaction) as far as their effect on capacity is concerned.

Regardless of the number of factors involved, certain strict assumptions must be considered in analysis of variance problems. They need to be discussed here, simply because they fail to hold in the setting of our problem. In the model of equation (3.1) it is assumed that the ε_{ij} are Gaussian with mean zero and common variance σ^2. These assumptions are necessary in order that the F-test used in the procedure be valid.

(b) Power of the Analysis of Variance

The analysis of variance procedure discussed here actually entails a test of the hypothesis that the means of each group, or the so-called "treatment means" do not differ from one another. Since the methodology falls into this traditional framework, the decision regarding the number of simulation runs, i.e., replications, the number of treatments or treatment combinations (number of conditions in our case) should be made with view toward obtaining a certain power of the test. The power of a test is defined as follows

$$P(\alpha,\nu_1,\nu_2,\Phi) = \Pr[\text{reject } H_0|H_1].$$

The power calculation is not a difficult one, in general, and if all of the above assumptions hold, involves the use of the non-central F distribution [3]. The distribution involves the following four parameters,

k = number of treatments

ν_1 = numerator degress of freedom = k-1

ν_2 = denominator degrees of freedom = N-k, where N is the total number of observations

$$\Phi = \sqrt{\frac{\sum_{i=1}^{k} n_i(\gamma_i-\bar{\gamma})^2}{k\sigma^2}} \tag{3.2}$$

The quantity Φ is called the non-centrality parameter. The power of the test increases monotonically with an increase in ν_1, ν_2, and of course the parameter Φ. As one can easily see, the test becomes more powerful or more sensitive to mean differences as the treatment means become more dispersed and as σ^2 becomes smaller, and as the number of replications becomes larger.

The most popular mode for finding the power of analysis of variance is the use of Operating Characteristic Curves or O.C. curves. Examples are displayed in Figures VII - X. The ordinate on the plots is actually the probability of making a Type II error, or the probability of false acceptance of H_0. Thus the power is found by calculating the complement of the number found on the abscissa. For example, let us assume that all standard conditions are met for analysis of variance and that there are 4 treatments to compare with 5 observations or replications per treatment. In addition, let us assume that the treatment means are $\gamma_1 = 55$, $\gamma_2 = 58$, $\gamma_3 = 61$, and $\gamma_4 = 64$. The parameter Φ is given by (assume $\sigma^2 = 25$)

$$\Phi = \sqrt{\frac{5(46)}{4(25)}} = 1.7$$

If one uses a significance level of the test of $\alpha = 0.05$, the power from the chart for $\nu_1 = 3$ and $\nu_2 = 16$ is found to be $1 - 0.3 = 0.7$. The implication is that if the real world situation provides a set of means given by the γ_i above, there will be 0.7 probability of rejecting H_0 or detecting a significant difference in the means.

In our application the data to be analyzed would be simulation data, and the γ_i would be mean throughput or capacity. In the following section we deal with the problem of failure of the analysis of variance assumptions.

IV. Failure of Analysis of Variance Assumptions

As we indicated earlier, normality of the ε_{ij}, i.e., normality of the treatment population is essential in order that the theory associated with the analysis of variance be intact. Also, the homogeneous variance assumption must hold. These assumptions allow one to use the F-test in analysis of variance. If they do not hold, analysis of variance is not correct in the sense that the distribution of the usual test statistic does not follow an F_{ν_1,ν_2}-distribution and thus an F would only be an approximation. That is, if one used an F, his true significance level would theoretically not exactly be what he thinks it is. The same, of course, would be true of the power since the validity of the non-central F relies on the same assumptions. The problem then becomes one of assessing how robust (or insensitive) the procedure of analysis of variance is to these assumptions. If there is a lack of robustness, then how can one adjust some feature of the total methodology in order to recapture validity.

It is known that analysis of variance is fairly robust to departures from the normal distribution assumption. However, there is no known work that investigates this robustness of analysis of variance to the distribution involved here, namely the inverse of a truncated normal. In the case of the homogeneous variance assumption, there is no effect on the significance level of the test since under the condition of H_0 (equal means) the variances in $\frac{1}{y} = \frac{1}{IAT}$ will be the same for each group (each treatment being identically distributed). However, when H_0 is not true, there is a lack of homogeneity of variance. Indeed, for $IAT \cap N(\mu_{IAT}, \sigma_{IAT})$, we found from II(a) that the approximate variance of throughput (on an hourly basis) is given by

$$\text{var}\ \frac{3600}{y} = \frac{3600^2 \sigma^2_{IAT}}{\mu^4_{IAT}}$$

where IAT is measured as sec/vehicle. Clearly, then, when one is testing the hypothesis that mean throughput (capacity) differs from condition to condition, he must cope with the fact that when H_1 is true, the variances also differ. Thus the computation of the power of the test cannot be made in the usual fashion because the non centrality parameter in (3.2) does not apply; indeed, there is no constant σ^2_{IAT} (σ^2_{IAT} here playing the role of Var $\frac{3600}{y}$). Thus, some adjustment in the computation of the power must be made.

(a) Simulation Study of Significance Level of Analysis of Variance

As was indicated earlier, there is a need to determine the adequacy of analysis of variance procedures when the basic measurement is throughput, $\frac{3600}{y}$. A simulation study was designed to determine how robust the significance level of the usual analysis of variance is to departures from the Gaussian assumption. There is no need to be concerned about heterogeneous variance here because the significance level α = Pr(reject $H_0|H_0$) would be studied by simulating distributions of $\frac{3600}{y}$ under the condition that the capacities are equal.

Various conditions on capacities, number of treatments, number of observations per treatment (number of touchdowns at that condition), etc. were fixed to use in the computer simulation. The $(\mu_{IAT}, \sigma_{IAT})$ combinations were controlled in order to establish the desired conditions on capacity. Basically, the model of equation (3.1) was used to generate data with each z_{ij} being $\frac{3600}{y}$ and $y \cap N(\mu_{IAT}, \sigma_{IAT})$. In order to provide a practical truncation of the normal random variable, the Monte Carlo Procedure truncated the normal variate at 51, i.e., effectively allowing no IAT below 51. A computer program was then written to compute the F-statistic for 2000 cases for each condition. The distribution of the F-statistic was then simulated and the probability of exceeding the "true" F percentage point was found. This then represents the significance level of the tests for analysis of variance in which the basic random variable is vehicle throughput. The results are given in Tables I and II. As one can easily see, the results strongly suggest that doing analysis of variance is quite good even when the basic measurement is not Gaussian but rather the inverse of a Guassian variate. The column headed "advertised α" is the significance level that one proceeds with in the analysis of variance methodology. The ± values indicate statistical fluctuation in the simulated significance level.

Table I

Simulated Significance Level

5 Treatments, 10 Observations per Treatment

Capacities	μ_{IAT}	σ_{IAT}	Advertised α	Actual α (Simulation)
60	60	5	0.05	0.0448±0.0024
51.4286	70	10	0.05	0.0516±0.0034
45	80	15	0.05	0.0494±0.00307
36	100	25	0.05	0.0496±0.00208
30	120	30	0.05	0.0434±0.00291
40	90	20	0.05	0.0528±0.00298
60	60	5	0.10	0.1005±0.0055
51.4286	70	10	0.10	0.0988±0.0022
45	80	15	0.10	0.099 ± 0.0033
40	90	20	0.10	0.1002±0.0045
30	120	30	0.10	0.0978±0.0038
60	60	5	0.20	0.1998±0.0040
51.4286	70	10	0.20	0.193 ± 0.0041
45	80	15	0.20	0.1876±0.0053
40	90	20	0.20	0.1988±0.0038
36	100	25	0.20	0.192 ± 0.00322
30	120	30	0.20	0.1984±0.0054
25.7143	140	45	0.05	0.048 ± 0.0035

Table II

Simulated Significance Level

4 Treatments, 10 Observations per Treatment

C	μ_{IAT}	σ_{IAT}	Advertised α	Actual α (Simulation)
20	180	65	0.05	0.0458±0.0029

2 Treatments, 10 Observations per Treatment

C	μ_{IAT}	σ_{IAT}	Advertised α	Actual α (Simulation)
20	180	65	0.05	0.0435±0.0085
20	180	65	0.10	0.0968±0.0031

(b) Simulation Study for Power of Analysis of Variance

Of course, knowing that the true significance level of the analysis of variance is very close to the one applied in the methodology is very comforting, but does not necessarily imply that one can generate a reasonable procedure for computing the power of the test; particularly since under H_1, i.e., when the treatments differ, varying values of μ imply heterogeneous variances. We would still like to be able to use the O.C. charts in Figures VII - X. The question then becomes "What is the effective value of σ^2 in the expression for the non centrality parameter in equation (3.2)?" However, when one fixes conditions on the capacities with varying μ and σ, this variance is not constant. As an adhoc methodology to apply in the calculation of Φ, consider the use of mean μ_{IAT} ($\bar{\mu}_{IAT}$), mean σ^2_{IAT} ($\bar{\sigma}^2_{IAT}$), and

$$\sigma_z^2 = \frac{(3600)^2\bar{\sigma}^2_{IAT}}{\bar{\mu}^4_{IAT}} ,$$

thus giving us for Φ, the relation 3.2 with σ^2 replaced by the "effective" σ_z^2 above. Again, simulation was done to study the appropriateness of this approximation of the power. Table III gives the results of this study.

The results in Table III are promising in that it does indicate a very definite procedure through the use of O.C. charts for approximating the power of the analysis of variance test even though the standard conditions of analysis of variance are not met. It should be emphasized that the above is not exhaustive and computer simulation costs would

Table III

Simulated and Computed Power of Analysis of Variance

Nominal Capacity ≅ 60,
4 treatments, 5 observations per treatment

Actual Capacities				μ_{IAT}				σ^2_{IAT}			
γ_1	γ_2	γ_3	γ_4	μ_1	μ_2	μ_3	μ_4	σ_1^2	σ_2^2	σ_3^2	σ_4^2
55	58	61	64	65.45	62.06	59.01	56.24	51.39	29.58	15.07	6.10

α	σ_z^2	Φ^2	Computed Power	Simulated Power
0.05	25	2.3	≅0.70	0.704 ± 0.0068
0.01	25	2.3	≅0.42	0.4368±0.0059

Same as Above, but 10 observations per treatment

α	σ_z^2	Φ^2	Computed Power	Simulated Power
0.05	15	4.6	≅0.98	0.9796±0.0055

Nominal Capacity ≅ 45,
5 treatments, 5 observations per treatment

Actual Capacities					μ_{IAT}					σ^2_{IAT}				
γ_1	γ_2	γ_3	γ_4	γ_5	μ_1	μ_2	μ_3	μ_4	μ_5	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
39	42	45	48	51	92.3	85.7	80	75	70.6	435.4	306.3	212.9	145	96

α	σ_z^2	Φ^2	Computed Power	Simulated Power
0.05	72	1.25	≅0.36	0.393 ± 0.0057

Same as Above, but 10 observations per treatment

α	σ_z^2	Φ^2	Computed Power	Simulated Power
0.05	72	2.5	≅0.75	0.7508±0.0047

Same as Above, but 20 observations per treatment

α	σ_z^2	Φ^2	Computed Power	Simulated Power
0.05	72	5.0	≅0.99	0.978 ± 0.006

Table III continued

Simulated and Computed Power of Analysis of Variance

Nominal Capacity ≅ 20
4 treatments, 5 observations per treatment

Actual Capacities				μ_{IAT}				σ^2_{IAT}			
γ_1	γ_2	γ_3	γ_4	μ_1	μ_2	μ_3	μ_4	σ^2_1	σ^2_2	σ^2_3	σ^2_4
25.7	22.5	18	16.4	140	160	200	200	2043.4	3070.1	5748.1	5748.1

α	σ^2_z	Computed Power	Simulated Power
0.05	53	≅0.28	0.2636±0.0049

Nominal Capacity ≅ 20,
2 treatments, 10 observations per treatment

Actual Capacities		μ_{IAT}		σ^2_{IAT}	
γ_1	γ_2	μ_1	μ_2	σ^2_1	σ^2_2
25.7	16.4	140	220	2043.1	7399.5

α	σ^2_z	Computed Power	Simulated Power
0.05	53	≅0.42	0.386 ± 0.0095

---

certainly not allow all types of relevant combinations. In addition, this is not intended as a study of the economic practicality of doing analysis of variance for detecting change in capacity, only as an assessment of a method for approximating power. A study of how costly such anova procedures would be (in terms of number of replications) will be the subject of the next section.

V. The Practical Use of Anova in Designed Simulation Experiments for Detecting Changes in Capacity

All of the preceeding results were designed to lead us to this section. Here we discuss the economic implication of designing simulation experiments to detect differences in capacity from one condition to the other. The basic idea is to get some notion of how much simulation effort (number of touchdowns) is necessary in order to have a successful simulation experiment. Of course the key here is to rely on the fact that we can conduct an analysis of variance and compute the power of the test on equality of capacity across conditions. We will assume at the outset that a significance level of $\alpha = 0.05$ would be used and that a successful test is one that would detect changes in capacity of 2-3 vehicles per hour with a high probability, say close to 0.95.

In other words, we shall display how many touchdowns per condition are required in order that one achieve a power of the test of 0.95 with a significance level of 0.05. It should be emphasized that this represents, by most standards, a very successful experiment. The procedure for computing the power involves the use of the O.C. charts and the computation of Φ as described in the previous section. In what follows we attempt to tabulate requirements on number of replications that result in a successful or near-successful experiment. These tabulations will be followed by some general conclusions. Table IV gives results on n (number of observations per treatment) for various values of nominal capacity, number of treatments, that give specified power.

Certain things become very obvious from the table. There is a more demanding sample size requirement for small capacities than for large capacities. The most demanding capacity is 30 (not 20) because σ_z^2 reaches its maximum at a value of approximately 30. In the case of comparing two treatments a fairly large number of replications is required in order to achieve the high power when nominal capacity is below 60. It would seem reasonable that for the case of two treatments, a higher significance level should be used in the test. Even the case of 3 treatments can require a fairly costly number of replications for low capacities while for four and five treatments, the number of replications required are not so demanding.

The above would suggest that the most efficient simulation experiment might involve the use of more than one factor. For example, two factors with 3 levels each (equivalent to six treatments) would be comparable to a study with 3 treatments and 3 n replications, where n is the number of observations in each combination. For experiments that must stay simple, say two or three treatments at low capacity, one must sacrifice something, perhaps using a somewhat higher significance level.

Table IV

Sample Size Requirements for Specified Power

Nominal Capacity	No. of Treatments	Significance Level	Capacities	n	Power
60	5	0.05	56,58,60,62,64	10	0.83
60	5	0.05	56,58,60,62,64	12	0.93
60	5	0.05	56,58,60,62,64	15	>0.95
60	4	0.05	55,58,61,64	10	>0.95
60	3	0.05	57,60,63	10	0.62
60	3	0.05	57,60,63	20	0.93
60	2	0.05	58,62	30	0.84
60	2	0.10	58,62	20	0.82
60	2	0.10	58,62	30	0.94
50	2	0.05	48,52	80	>0.95
50	2	0.10	48,52	70	>0.95
45	5	0.05	36,38,40,42,44	20	>0.95
45	4	0.05	40,43,46,49	25	0.95
45	3	0.05	42,45,48	60	0.95
30	5	0.05	26,28,30,32,34	30	0.9
30	5	0.05	26,28,30,32,34	35	>0.95
30	4	0.05	25.5,28.5,31.5,34.5	30	0.93
30	4	0.05	25.5,28.5,31.5,34.5	35	>0.95
30	3	0.05	27,30,33	60	0.93
20	5	0.05	16,18,20,22,24	20	0.85
20	5	0.05	16,18,20,22,24	25	0.94
20	4	0.05	15.5,18.5,21.5,24.5	20	0.92
20	4	0.05	15.5,18.5,21.5,24.5	25	0.99
20	3	0.05	17,20,23	50	0.95
20	2	0.05	18,22	70	0.91
20	2	0.05	18,22	80	0.95

In this section we illustrate the foregoing methodology with three examples that represent actual hypothetical simulation experiments that might be of interest to NASA Langley personnel. (See Ref. 2).

example 1

Suppose it is of interest to compare two controller conditions, condition A being M&S (metering and spacing) and condition B being vectoring. Suppose B has a capacity of 38.6 vehicles per hour, resulting from μ_{IAT} = 93.3 seconds and σ_{IAT} = 18 seconds with Z_B representing a 1% Gaussian Buffer. Condition A has a capacity of 46.7 vehicles per hour resulting from μ_{IAT} = 77.1 seconds and σ_{IAT} = 11 seconds and Z_B represents a 1% Gaussian Buffer. Both A and B have s = 2 miles. In calculation of the value of n, the number of arrivals required to give adequate power, we have

$$\sigma_Z^2 = \frac{(3600)^2\bar{\sigma}^2_{IAT}}{\bar{\mu}^4_{IAT}} = 51.803$$

and thus

$$\Phi = \sqrt{\frac{[(38.6-42.65)^2+(46.7-42.65)^2]n}{2(51.803)}}$$

$$= \sqrt{\frac{(32.805)n}{103.6}}$$

Using the O.C. Charts in Figure X gives n=24 arrivals required per condition in order to achieve a power of 0.95 with a significance level of 0.05. This is required in order that there be a good probability of detecting differences between capacity at saturation when in fact the stated conditions A and B exist. These 24 arrivals will require approximately 0.6 hours per condition.

<u>example 2</u>

Consider two conditions A and B with both representing vectoring conditions with s = 2 miles for condition A and s = 3 for condition B. Allow Z_B to represent a 1% buffer for condition A and 5% for condition B. For condition A we have a capacity of 38.6 vehicles per hour with μ_{IAT} = 93.3 seconds, σ_{IAT} = 18 seconds, while conditions B has a capacity of 33.7 vehicles per hour, μ_{IAT} = 106.75 seconds, and σ_{IAT} = 18 seconds. Using the same approach as illustrated in example 1, we require n = 47 in order to achieve a value of $\Phi = 2.6$ required to achieve a power of 0.95 with a significance level of 0.05. This value of 47 arrivals requires approximately 1.3 hours per condition.

<u>example 3</u>

Let condition A represent metering and spacing with a capacity of 46.7 vehicles per second, μ_{IAT} = 77.1 seconds, σ_{IAT} = 11 seconds with a 1% buffer. Let condition B represent Cockpit Display of Traffic Information (CDTI), with a capacity of 52.9 vehicles per hour, μ_{IAT} = 68 seconds, σ_{IAT} = 3 seconds, representing a near 0% buffer. s = 2 miles for each condition. A value of n = 27 arrivals produces $\Phi = 2.65$, required for a power of 0.95. As a result, 0.6 hours would be required for each condition.

VII. Conclusions

We have demonstrated that one can effectively use analysis of variance to do significance testing in simulation experiments where vehicle throughput is the basic response. The difficulty with the failure of the normality assumption presents no serious problem in conducting the statistical tests. In computing the power of the tests, one needs to make an adjustment in the computation of the non centrality parameter Φ in using standard operating characteristic curves. The adjustment results in an approximate power and takes into account the heterogenous variances that are implicit in testing for differences in capacity between various conditions.

For designing experiments that compare capacities, it turns out that in order to achieve an exceedingly high power for moderate changes in capacity (2-4 vehicles per hour), a moderate number of replications are needed when at least four or five "treatment combinations" are to be studied in the design. For only two treatments, the number of replications are probably prohibitive, except for very high capacities.

Bibliography

1 Kendall, M.G. and A. Stuart. *The Advanced Theory of Statistics*, Vol. I, Hafner Publishing Co., 1969.

2 Swedish, W.J. "Evaluation of the Potential for Reduced Longitudinal Spacing on Final Approach," The Mitre Corporation, Metrek Division, MTR-79W00280 (FAA-EM-79-7), August, 1979.

3 Walpole, R.E. and Raymond H. Myers. *Probability and Statistics for Engineers and Scientists*, MacMillan Co., 1972.

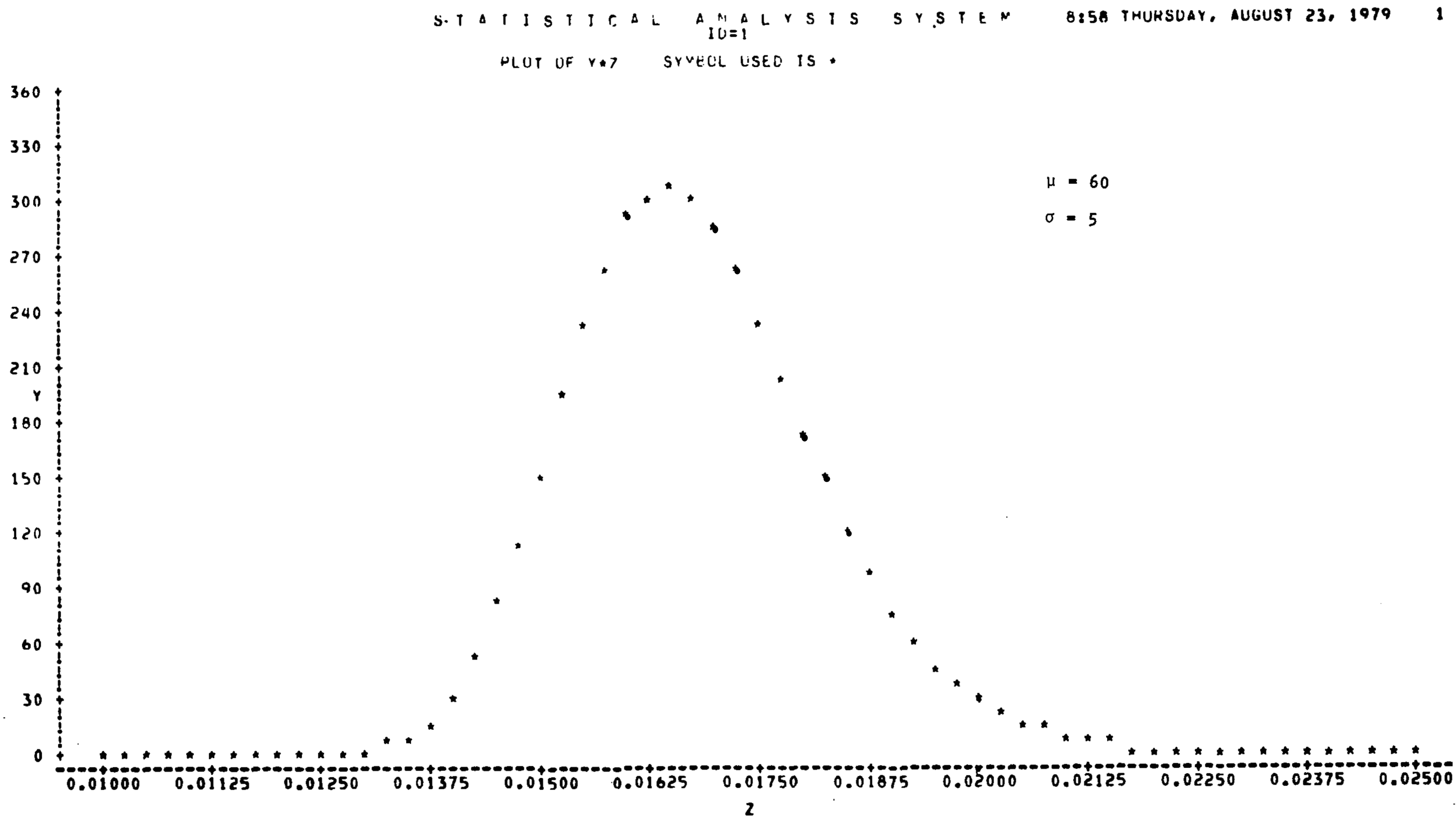


FIGURE I. Distribution of Z

STATISTICAL ANALYSIS SYSTEM 8:58 THURSDAY, AUGUST 23, 1979 2

ID=2

PLOT OF Y*Z SYMBOL USED IS *

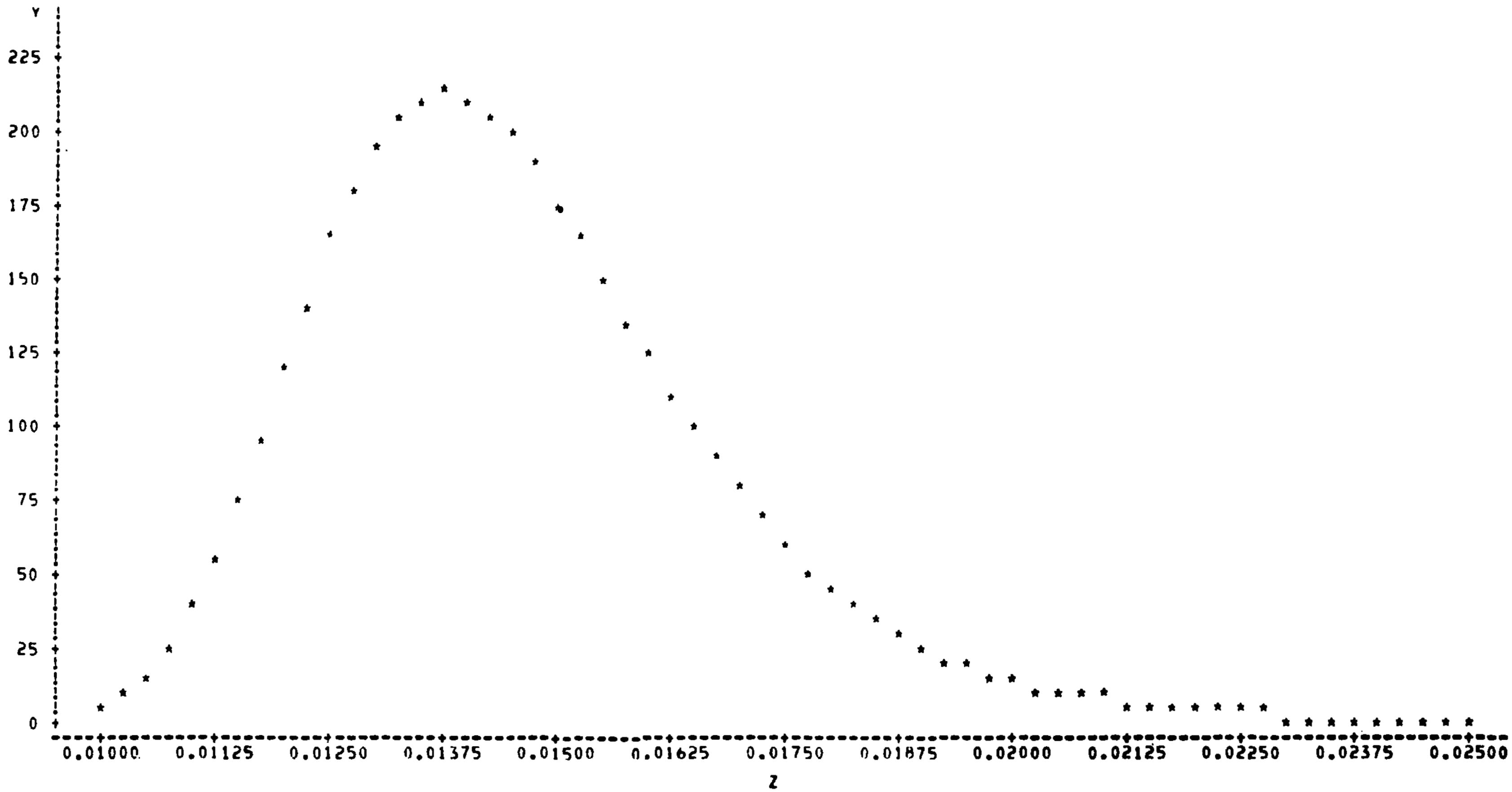


FIGURE II. Distribution of Z

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   8:58 THURSDAY, AUGUST 23, 1979   3

ID=3

PLOT OF Y*Z   SYMBOL USED IS *

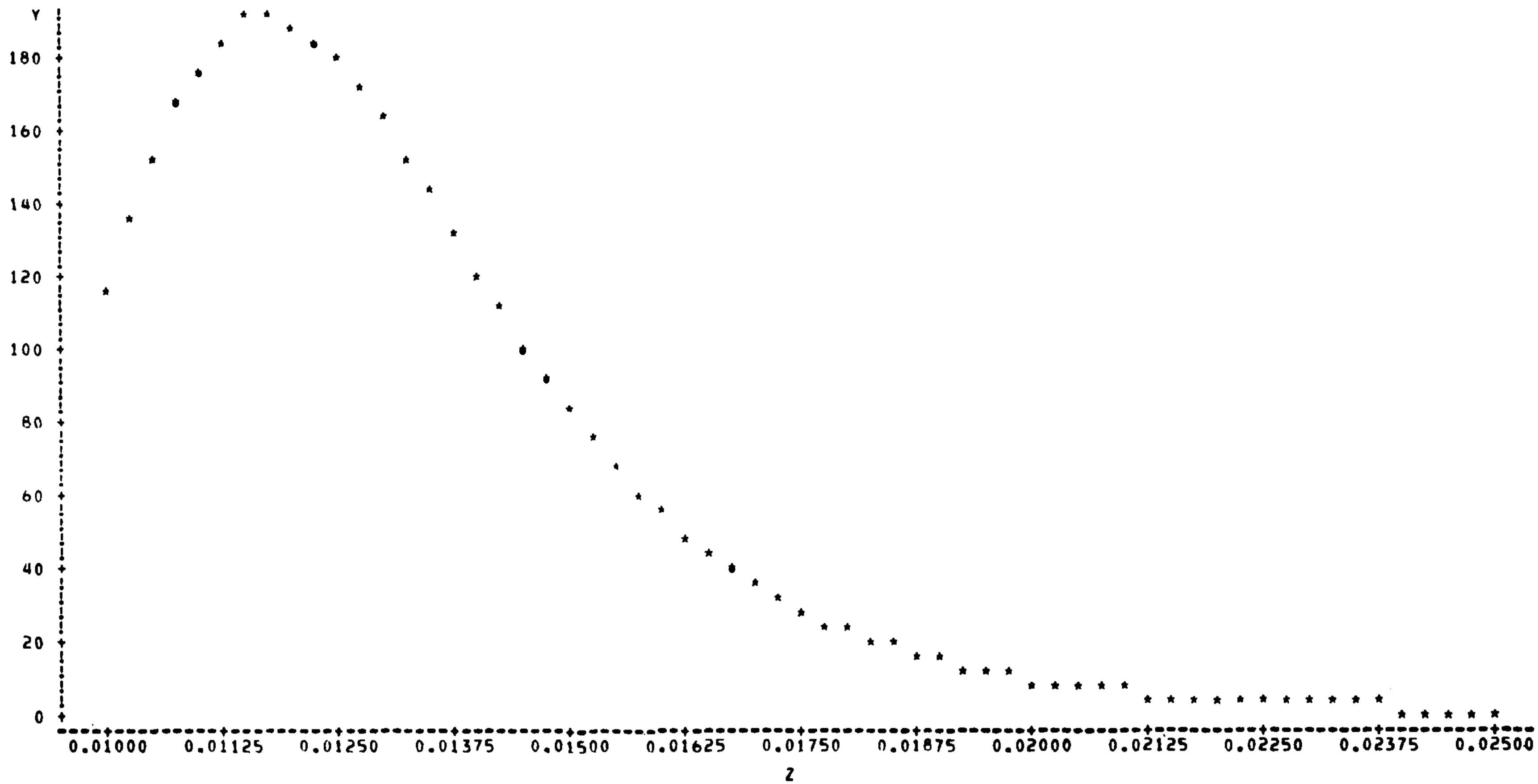


FIGURE III. Distribution of Z

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   6:56 THURSDAY, AUGUST 23, 1979   4

ID=4

PLOT OF Y*Z   SYMBOL USED IS *

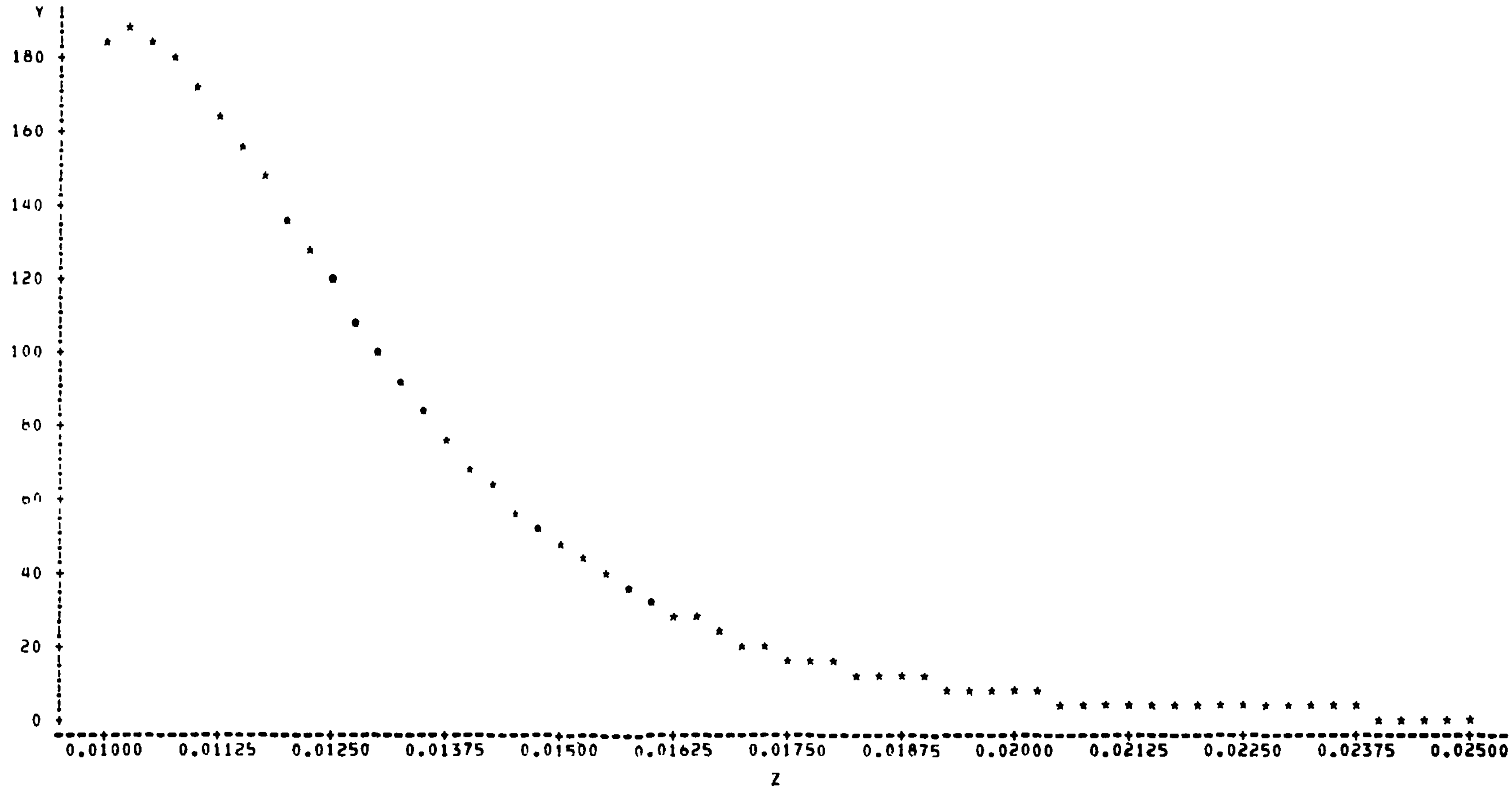


FIGURE IV. Distribution of Z

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   8:58 THURSDAY, AUGUST 23, 1979   5
ID=5

PLOT OF Y*Z   SYMBOL USED IS *

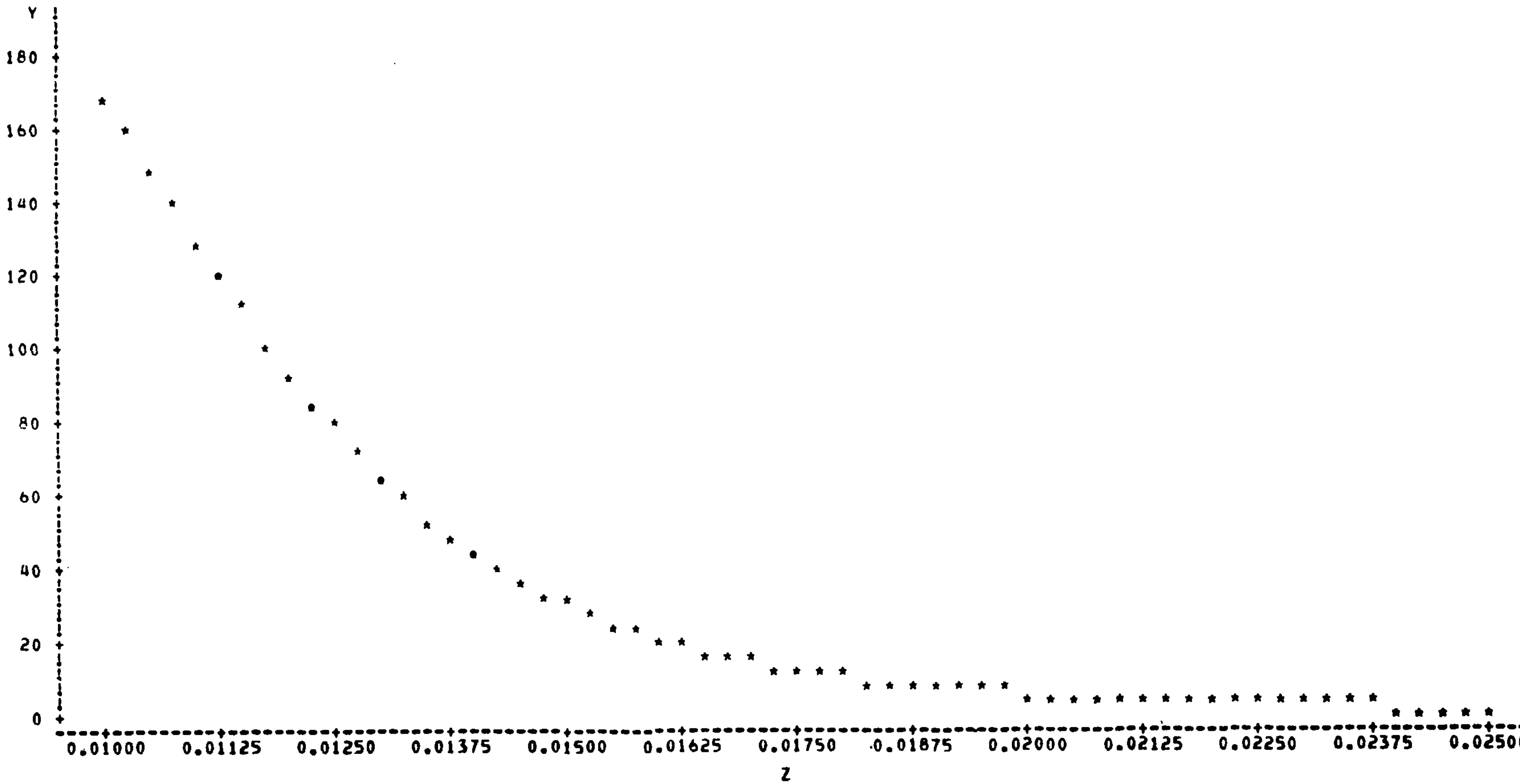


FIGURE V. Distribution of Z

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   8:58 THURSDAY, AUGUST 23, 1979   6
ID=6

PLOT OF Y*Z   SYMBOL USED IS *

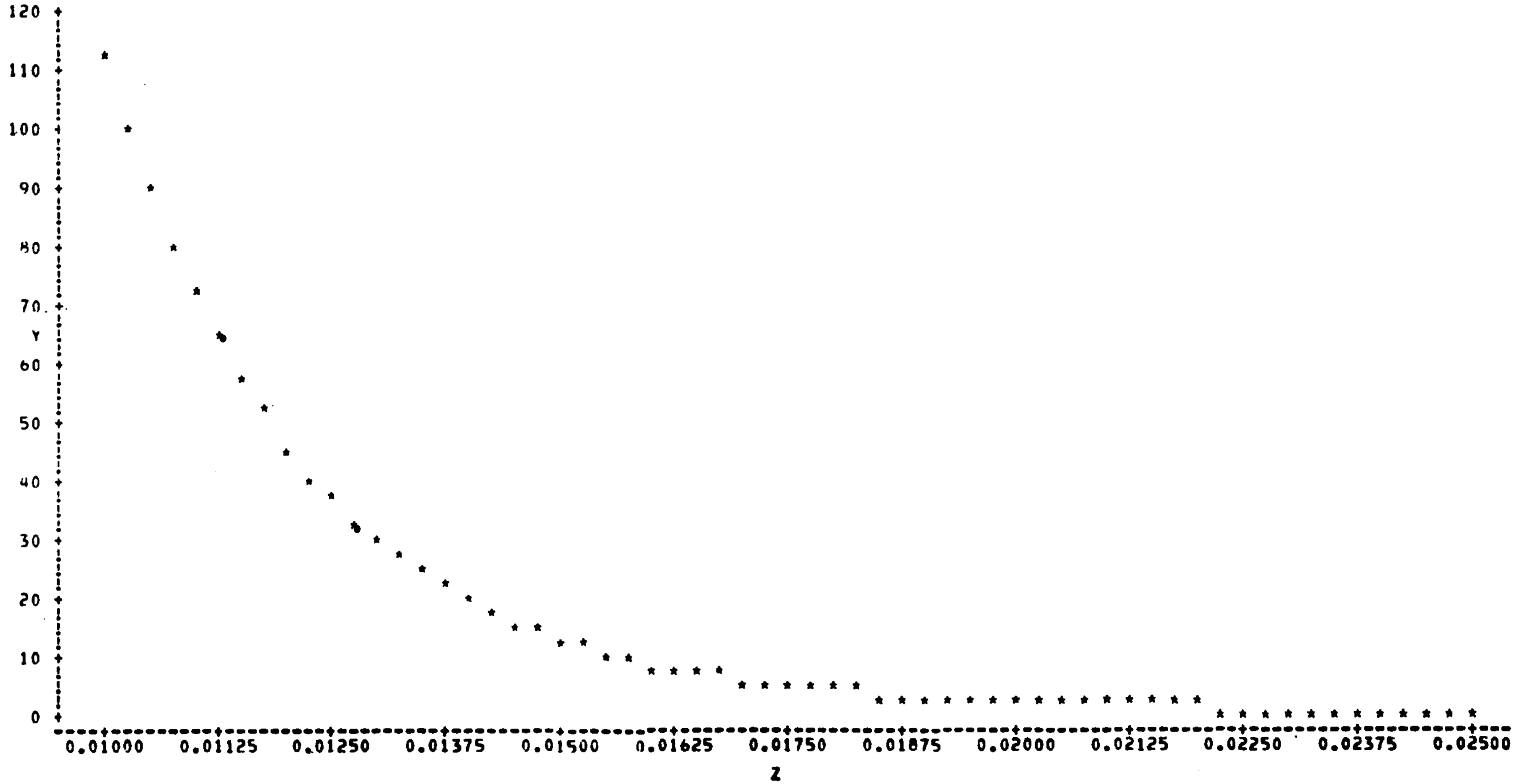
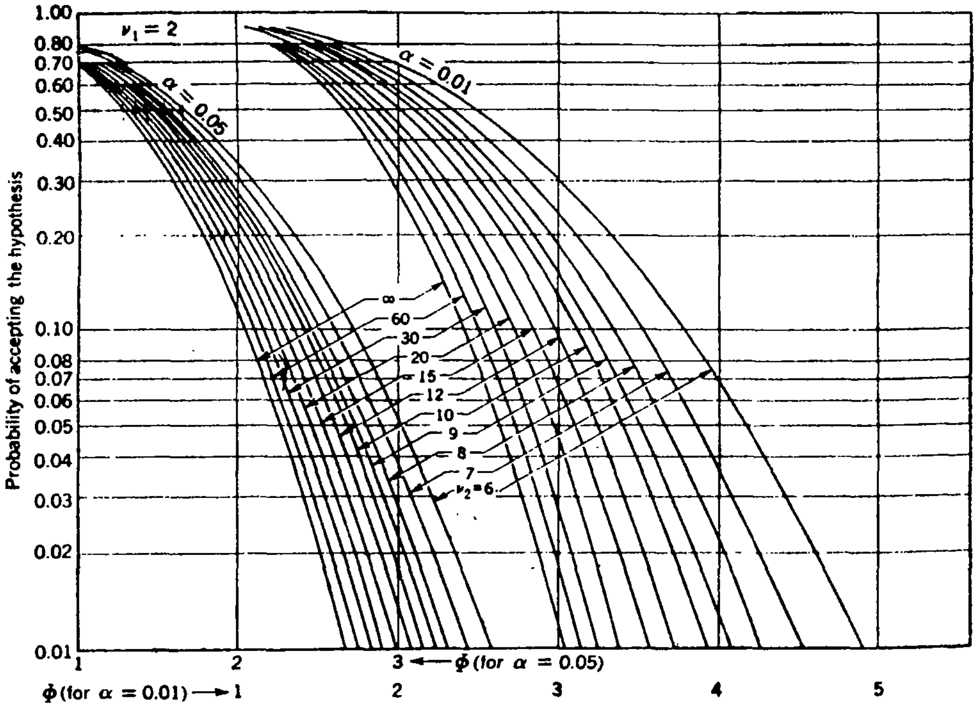
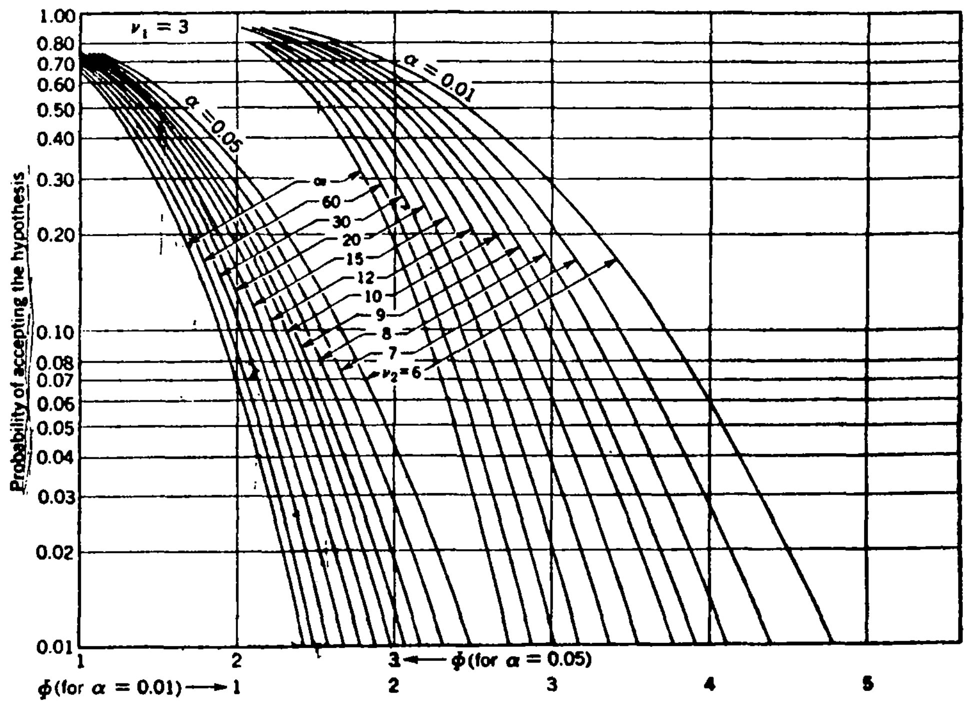


FIGURE VI. Distribution of Z

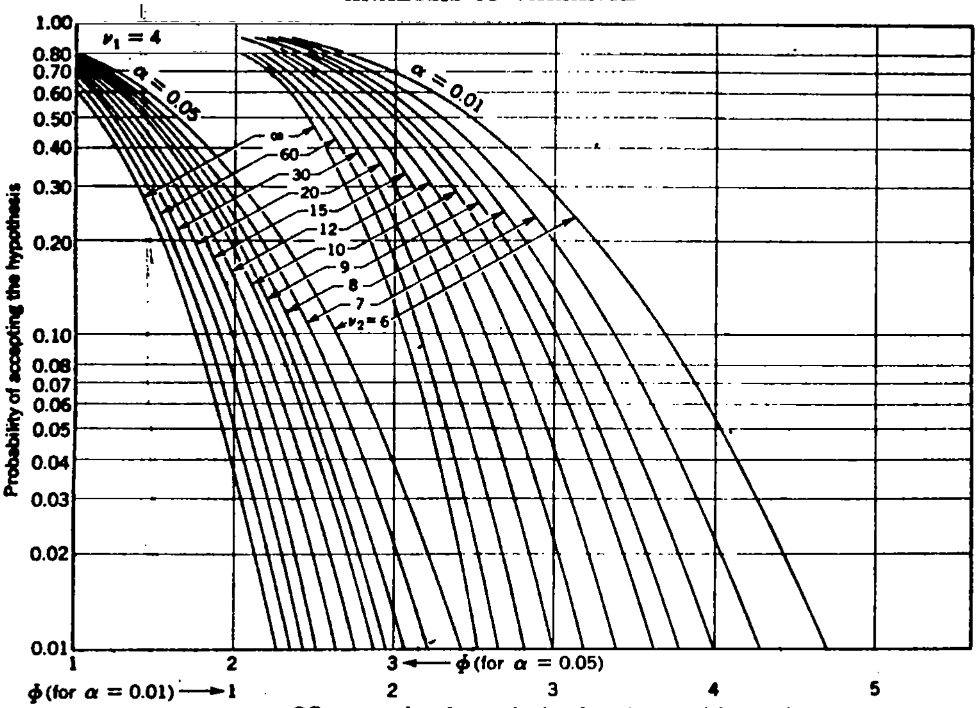


OC curves for the analysis of variance with $\nu = 2$.

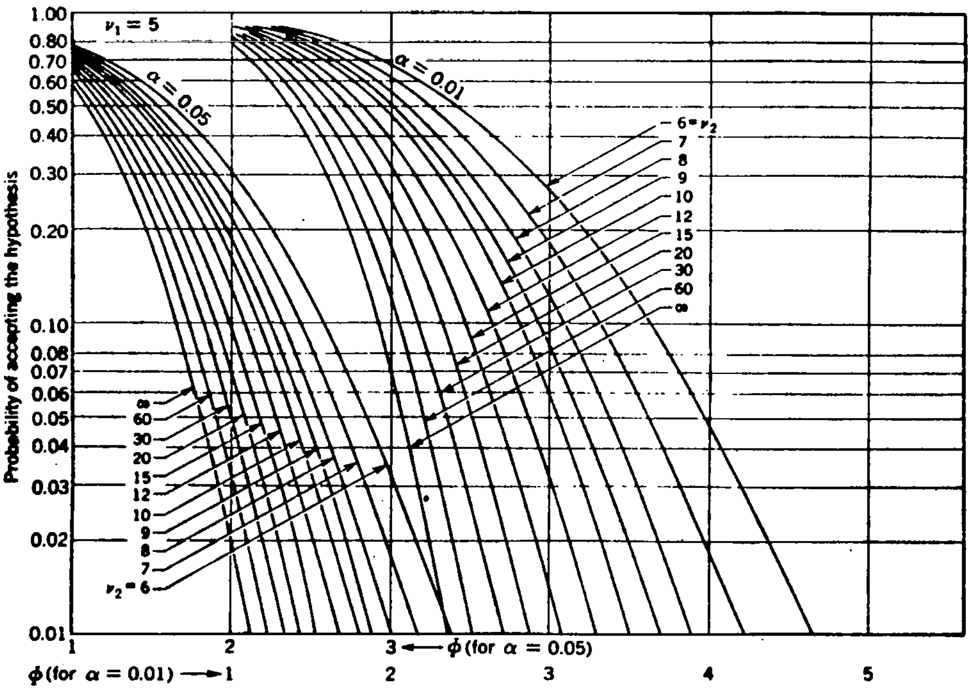


OC curves for the analysis of variance with $\nu = 3$.

FIGURE VII

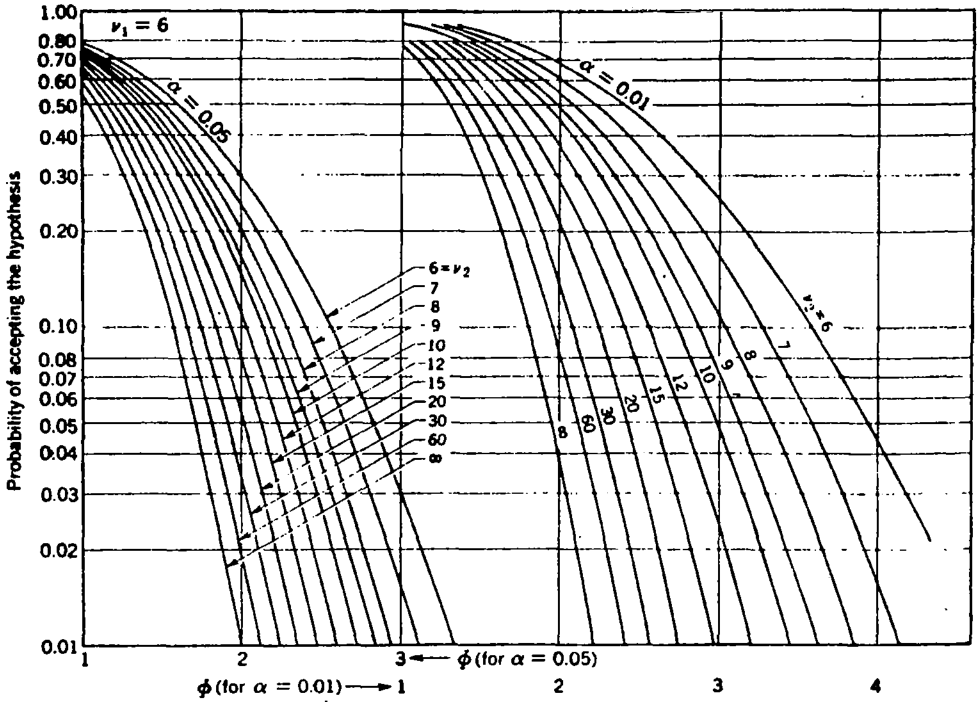


OC curves for the analysis of variance with $\nu = 4$.

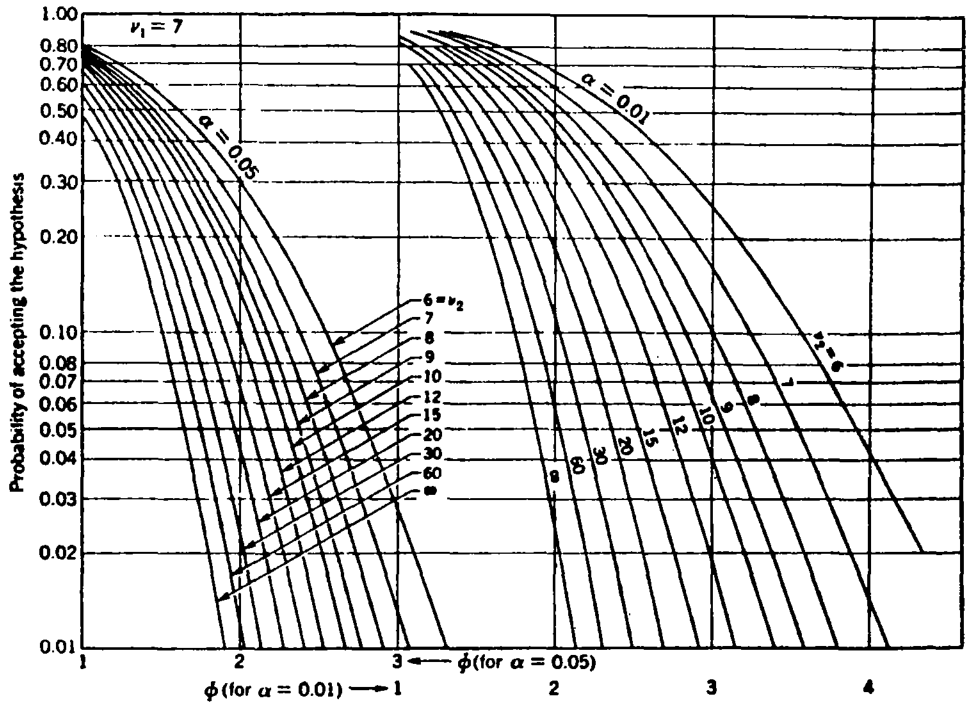


OC curves for the analysis of variance with $\nu = 5$.

FIGURE VIII

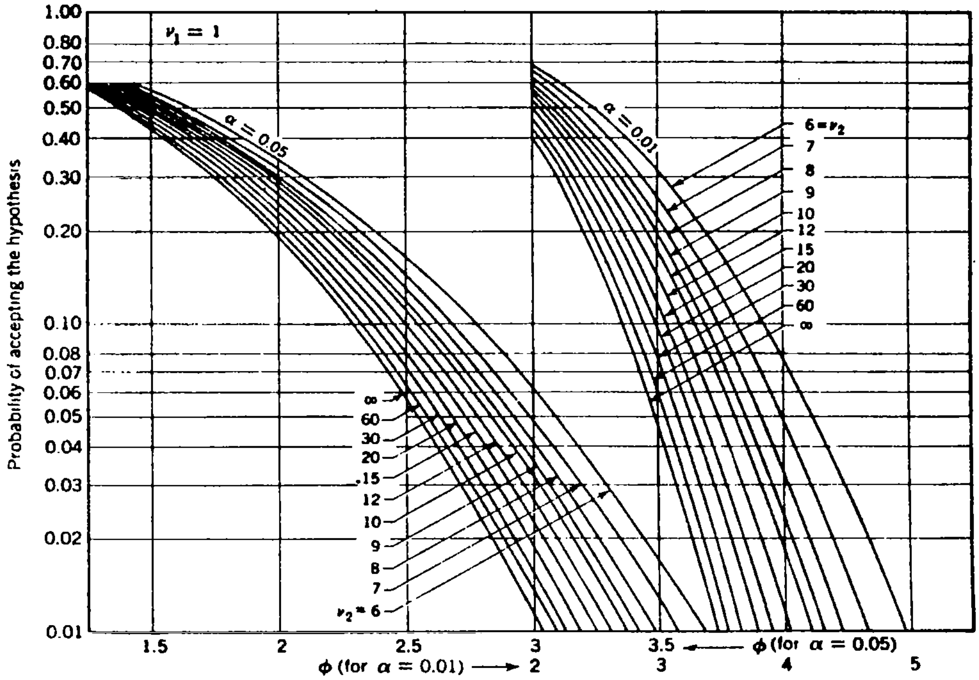


OC curves for the analysis of variance with $\nu = 6$.



OC curves for the analysis of variance with $\nu = 7$.

FIGURE IX



OC curves for the analysis of variance with $\nu = 1$.

FIGURE X

1. Report No. NASA CR-3633	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A STUDY OF THE FEASIBILITY OF STATISTICAL ANALYSIS OF AIRPORT PERFORMANCE SIMULATION		5. Report Date November 1982
		6. Performing Organization Code
7. Author(s) Raymond H. Myers		8. Performing Organization Report No.
		10. Work Unit No.
9. Performing Organization Name and Address RAYMOND H. MYERS 206 FINCASTLE DRIVE BLACKSBURG, VA 24060		11. Contract or Grant No.
		13. Type of Report and Period Covered Contractor Report
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS & SPACE ADMINISTRATION WASHINGTON, DC 20546		14. Sponsoring Agency Code 505-35-33-02 (L93890A)

15. Supplementary Notes

LANGLEY TECHNICAL MONITOR: BURNELL T. McKISSICK
FINAL REPORT

16. Abstract

Any proposed simulation study of airport capacity by NASA and FAA should be carried out with the aid of sound statistical tools. This study is a fundamental one in which the feasibility of conducting statistical analysis of simulation experiments to study airport capacity is investigated.

First, a study of the form of the distribution of airport capacity is made. The distribution is non-Gaussian, it is important to first determine the effect of this distribution on standard analysis of variance techniques and power calculations. Next, power computations are made in order to determine how economic simulation experiments would be if they are designed to detect capacity changes from condition to condition. Many of the conclusions drawn here are results of Monte-Carlo Techniques.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Simulation Experiments, Airport Capacity Non-Gaussian Distribution, Analysis of Variance, Power Calculations, Monte-Carlo Techniques.	Unclassified-Unlimited Subject Category 65

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	36	A03

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1982