STATISTICAL THEORY OF SELECTION (SELECTION PROBLEMS).
SINGH, JARNAIL
ProQuest Dissertations and Theses; 1963; ProQuest Dissertations & Theses Global

This dissertation has been microfilmed exactly as received

64–7121

SINGH, Jarnail, 1932–
STATISTICAL THEORY OF SELECTION (SELECTION PROBLEMS).

University of Toronto, Ph.D., 1963
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

STATISTICAL THEORY OF SELECTION

(Selection Problems)

by

JARNAIL SINGH

A Thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy

in

The University of TORONTO

1963

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UNIVERSITY OF TORONTO

SCHOOL OF GRADUATE STUDIES

PROGRAMME OF THE FINAL ORAL EXAMINATION

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

OF

JARNAIL SINGH

3:30 p.m., Friday, October 25th, 1963

School of Graduate Studies

STATISTICAL THEORY OF SELECTION
(Selection Problems)

Committee in Charge:

Professor I. H. Spinner, Chairman
Professor D. B. De Lury
Professor D. A. S. Fraser
Professor R. Wormleighton
Professor C. C. Gotlieb
Professor W. Knight
Professor J. I. McDonald
Professor P. G. Rooney
Professor M. Srivastava
Professor W. Webber
Professor Beatrice Worsley

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

THESIS

STATISTICAL THEORY OF SELECTION
(Selection Problems)

(Summary)

The twin problems of selection and screening of varieties have been considered. In selection we are required to select the best or 't' best out of k (k> t) varieties. The selection criterion is the sample mean of the variety. Each observation on a variety is independent and follows normal distribution with known variance and unknown mean characteristic of the variety. We further assume that the varietal means themselves follow some distribution. Three distributions, exponential, gamma and double exponential have been investigated. For the exponential and the gamma probability of correct selection, i.e., the probability that the best variety is selected, conditional probability of correct selection and the probability that the selected variety is within δ (specified) of the best variety have been evaluated. For the double exponential, only the first two of the three probability formulations have been considered.

For the exponential distribution an alternative procedure for the probability of correct selection is given. From that we get that the probability of the rth best variety exceeding the best one is independent of k (r<k)

Numerical results have been compared with the normal distribution. It appears that for k=2 the probability of correct selection does not depend very much on the a priori distribution of the varietal means. A procedure is given by which the probability of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

correct selection, provided the error variance is such that the probability value will be of the order of 90%, can be obtained for all k for the exponential distribution. The same procedure has been adopted for double exponential and normal distributions to extend the numerical results up to k=4 and k=3 respectively.

A number of miscellaneous problems, such as optimum size of the experiment, selection of the 't' best varieties, complete ranking of the varieties are considered. Finally, selection based upon two characters is considered.

In screening we are interested in all varieties which are above a certain level. The procedure adopted is to select those varieties which are better than the control by a specified amount. Three distributions, normal, exponential and double exponential have been considered. Screening results in an increase in the mean yielding capacity of the selected varieties over the whole cohort. The increase in the yielding capacity has been expressed in terms of the standard deviation of the distribution in question and is termed standard gain. The fraction selected for low values of $\left(\frac{\sigma_0}{E}\right)$, ($\sigma_0$ is the standard deviation of the a priori distribution and $E=\sqrt{(2/n)}\sigma$; σ is the error variance) is almost the same for the three distributions for the first stage selection, but the standard gain is higher for the exponential. The gain is the expected gain; an expression for the variance of standard gain is obtained, it depends upon the replications and the number of varieties tested.

In the two stage screening two alternative procedures, i.e., considering the first stage results and not considering the first stage results are separately dealt with. Here again the exponential distribution gives higher standard gain. Finally, screening based upon two characters is considered.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

BIOGRAPHY

1932 -	Born, V. Sirhala, Lyallpur (Now Pakistan)
1952 -	B.Sc.Agri. Khalsa College, Amritsar, Panjab, India
1954 -	M.A. Econ. Khalsa College, Amritsar, Panjab, India
1961 -	M.A. University of Toronto
1961 - 1963	Teaching Fellow, Department of Mathematics, University of Toronto
1959 - 1963	School of Graduate Studies University of Toronto

GRADUATE STUDIES

Major Subjects:

Design of Experiments	Professor D. B. DeLury
Linear Statistical Analysis	Professor G. S. Watson Professor R. Wormleighton

Minor Subjects:

Statistical Inference	Professor D.A.S. Fraser
Mathematical Economics	Professor J. I. Macdonald

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Acknowledgments

I wish to express my sincere gratitude to:

Prof. R. Wormleighton, under whose direction the present work has been carried out, for his expert guidance and keen interest in my work, for reading the preliminary draft, making valuable comments, which helped to clarify a number of points.

Mrs. H.M. Mann of O.A.C. Guelph for her typing.

The National Research Council of Canada for financial grants.

Jarnail Singh

September 1963.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Statistical Theory of Selection

(Selection problems)

An abstract of a Thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy in the University of TORONTO.

It is rather difficult to give technically correct definition of a variety. As a working definition, by variety we mean a separate and distinguishable type of grain, alloy, drug, or a strain of bacteria, as the case may be.

In this work the twin problems of selection and screening of varieties are considered. In selection, we are given a number (k) of varieties and we are required to select the best or 't' ($t < k$) best varieties. On the other hand, in screening, we are interested in all the varieties which are above a certain level or in some cases we may be interested in retaining a certain specified fraction of the population of the varieties.

The traditional analysis of variance procedure is not very suitable for selection problems, because when the experimenter knows a priori that the varieties are different there is no purpose in testing the hypothesis of homogeniety. The criterion for selection is the sample yield, and yield is supposed to be a continuous variable. On each variety a number of observations are taken; the number

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of observations is the same for all varieties. The variety which gives the maximum sample mean is taken to be the best.

The formulation of the selection problem will depend upon the a priori information available and the purpose in hand. In the absence of any information about the means of the varieties it may be desirable to formulate the problem in terms of least favourable configuration. The least favourable configuration expression is used for the situation when all the varieties are just δ (specified) inferior to the best variety. If some information a priori is available it should be utilized.

We assume that each observation on a variety is normal with unknown mean and known variance and the varietal means themselves follow some distribution. Three distributions, exponential, gamma and double exponential are investigated. Three formulations of the selection problem are considered, (a) probability of correct selection under all circumstances, i.e., the probability that the best variety is selected, (b) probability of correct selection under the condition that the best variety exceeds all others by a certain specified δ, and (c) the probability that the selected variety is within δ of the best. Numerical results are obtained for $k = 2$ for the probability of correct selection and those have been compared with normal distribution and least favourable configuration results. The least favourable configuration results are obtained by taking the difference of the two means

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

equal to the expected range of the two variables following the distribution under study, and it gives higher probability values. The normal distribution gives higher values than the other distributions but, the difference is not much. It appears that for $k = 2$ the probability of correct selection does not depend much upon the distribution and this is particularly so if the probability of correct selection is high.

For the exponential distribution an alternative formulation of the probability of correct selection is given; from that the interesting result we get is that the probability of the r-th best variety exceeding the other varieties is independent of k ($r < k$).

A procedure is given by which the probability of correct selection for the exponential distribution can be obtained for all k, provided the error variance is such that the probability of correct selection is of the order of 90%. The same method has been adopted to get the numerical values for k up to 4 in case of double exponential and for $k = 3$ in case of normal.

A number of miscellaneous selection problems, such as optimum size of the experiment when the varietal means follow exponential distribution; selecting of 't' best varieties and complete ranking of the varieties, are considered. Finally, selection based upon two characters is considered.

The second part of the Thesis deals with screening problem. Screening is in comparison with a standard, i.e., only

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

those varieties are selected which exceed the standard by a certain specified amount. Firstly, we deal with first stage screening. Three distributions, normal, exponential and double exponential are considered. Our object in screening is to have an increase in the mean yielding capacity of the selected varieties over the whole population. The gain in the mean yielding capacity expressed in units of the standard deviation of the distribution under study has been termed as standard gain. For the three distributions mathematical formulae are derived for the fraction selected and the standard gain obtained; numerical results are obtained for the same. The general inference from the numerical results we get is that if $\frac{\sigma_o}{E}$ (σ_o is the standard deviation of the parent population, $E = \sqrt{\frac{2}{n}}\,\sigma$; σ is the error variance) is not very large (< 1.0) and cut-off point is such that 1/10 or less of the population is selected, the fraction selected is almost the same for the three distributions. However, the standard gain is higher in the exponential distribution. The gain is the expected gain; an expression for the variance of the standard gain is obtained; it depends upon the error variance and the number of varieties tested.

In the two stage screening, two alternative procedures, e.g., considering the first stage result and without considering the first stage result are separately deal with. Mathematical formulae are derived for the fraction selected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the standard gain for the three distributions but numerical results are obtained for the normal and the exponential distributions only. Here again exponential gives higher standard gain. Finally, screening based upon two characters is considered; only mathematical formulae are derived.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table of Contents

		Page
	Abstract	iv-viii
Chapter 1.	Introduction, Review of the Literature and Statement of the problems	1-22
	Part I Selection	
Chapter 2.	Exponential Distribution	23-46
	I. Probability of Correct Selection	23
	II. Conditional Probability of Correct Selection	30
	III. Probability that the selected Variety is within δ of the Best	35
	IV. An Alternative Approach for Probability of Correct Selection	39
Chapter 3.	Gamma and Double Exponential Distributions	47-79
3.1	Gamma	47-61
	I. Probability of Correct Selection	47
	II. Conditional Probability of Correct Selection	53
	III. Probability that the Selected Variety is within δ of the Best	57
3.2	Double Exponential	61-79
	IV. Probability of Correct Selection	62
	V. Conditional Probability of Correct Selection	73
Chapter 4.	Numerical Evaluations	80-93
	I. Exponential	80
	II. Double Exponential	86
	III. Normal	88
	IV. Exponential - Conditional Probability	92

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5.	Miscellaneous Extensions		94-116
	I.1	Optimum Size of the Experiment	98
	II.	A Variation of the Probability of Correct Selection Problem	102
	III.	Selection of 't' Best Varieties	106
	IV.	Complete Ranking	112
Chapter 6.	Selection on Two Characters		117-128
	Part II Screening Problem		
Chapter 7.	First Stage Screening		129-158
	I.	Screening Problem	129
	II.1	Normal	137
	II.2	Exponential	139
	II.3	Double Exponential	140
	III.	Numerical Results	141
	IV.	Variance of Gain	149
	V.	Increased Number of Replications for the Standard	155
	VI.	Higher Moments of the Population After First Stage Selection	156
Chapter 8.	Second Stage Screening		159-179
	8.1	Without Considering the First Stage Result	159-168
	I.	Normal	159
	II.	Exponential	162
	III.	Double Exponential	165
	8.2	Considering the First Stage Result	169
	IV.	Normal	169
	IV.1	Exponential	170
	IV.2	Double Exponential	173

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 9. Screening on Two Characters 180-189

I. Introduction 180

II. First Stage 181

III. Second Stage 184

IV. Exponential 189

Glossary of Symbols 190

Bibliography 191-197

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1

INTRODUCTION, REVIEW OF THE LITERATURE AND STATEMENT OF THE PROBLEMS

Introduction

In development work such as evolving of new varieties of grain, new strains of bacteria, new types of alloys or new types of drugs - hereafter all called varieties - it is the endeavour of the experimenter to evolve new and better varieties than the existing ones. Towards this end the first step is the production of new varieties. Following preliminary culling the new varieties will be put to test in comparison with the standard/s. If any of the new varieties proves to be sufficiently superior to the existing standard its stock will be multiplied for commercial purposes. The production of new varieties is the job of the geneticist, botanist, chemist, or the engineer as the case may be and as such is not a statistical problem.

The selection is to be based on the performance of the varieties in respect to a single variate. The variate may be termed as yield. The way 'yield' is defined will depend upon the particular problem in hand. For agricultural crops it will be the quantity of grain or roots per unit area, for antibiotic producing micro-organisms it will be the amount of antibiotic produced per unit material consumed etc. However it is assumed that the variate-yield-is a continuous variable. Selection can be based on more than one character; for example in varieties of grain selection could be based

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

on quantity of grain, resistance to disease, drought resistance, or bread making quality of the grain. Similarly, in case of drugs, besides the main effect under consideration, selection may be based on the side effects on the host.

The traditional analysis of variance is not very suitable for this problem. It answers only one question that is if the varieties are homogenous in respect to the variate or not. But such a hypothesis is of very little value when the experimenter knows a priori that they are different. In newly produced strains it is taken for granted that they are different in their yielding capacities, otherwise there is no purpose to the whole breeding programme. Hence when the number of varieties is moderately high, say of the order of 15 to 20, the analysis of variance is almost sure to give some significant results. In this case the classical notion of significance is not relevant because even if the significance of the differences cannot be established the varieties giving higher (or lower) yield will be selected over the others. In case the decision is to be based on more than one character the analysis of variance does not afford much guidance. It is possible to analyse the two variates separately but that does not lend itself to any unified and coherent decision procedure. Alternative to the analysis of variance are the selection and screening formulations.

In "selection" we may include ranking and slippage

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

formulations. Selecting the best out of K varieties is a particular case of the ranking problem. The sample size required to ensure a specified probability of correct selection will naturally depend upon the relative distance of the means of the varieties. In the extreme case, we may ~~amy~~ not know any thing about the means. In such a case formulation of the problem as least favourable configuration may be desirable. However, if some information a priori is available it will be worthwhile to utilize that information. In the first part we develop such a procedure when the means themselves are supposed to follow certain distributions.

In chapter 2 the exponential and in chapter 3 double exponential and gamma distributions are assumed as a priori distributions and the numerical results for K = 2 are compared with the normal distribution as worked out by Dunnett (19) and with least favourable configuration as worked out by Bechhoffer (5).

Chapter 4 deals with the evaluation of the numerical results for K greater than 2. In the exponential case when the error variance is such that the probability of correct selection is greater than 0.90, we have obtained numerical results for all K. An approximate method is given by which we can get numerical results for double exponential for K 3 and 4. The same method has been adopted for the normal to get the numerical result for K = 3. In chapter 5 related formulations of the problem, such as optimum size of the experiment, complete ranking and selection of more than

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

one variety, etc., are discussed. Chapter 6 deals with the problem when the selection is based on two characters.

The second part deals with the screening problem. Selecting the best or the 't' best varieties is the last stage of the whole breeding programme. The stages of the programme prior to the final selection of the best variety may be called screening. In practice every newly produced cohort will contain a large proportion of varieties which should be dropped at the earliest possible. The object of screening is to skim off relatively superior varieties with minimum effort. The procedure adopted is to select only those varieties which exceed the standard by a specified amount. Chapter 7 deals with the first stage only and an estimate of the variance of the expected gain in the yielding capacity of the selected varieties over the whole cohort is given.

In chapter 8 consequences of a two stage screening programme are evaluated. Two alternatives, i.e. considering the first stage result and not considering the first stage result at the second stage are separately dealt with. Numerical results giving the percentage of the population selected and the expected gain have been obtained for the exponential, double exponential and the normal a priori distributions. Finally, in chapter 9, screening as based on two characters is considered.

Review of the Literature

Selection problems have been formulated in various ways.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The various formulations are attempts to have optimum procedures for selection under different conditions. Basically they correspond to the two categories of design of experiments.

(a) corresponding to model I, the varieties are fixed, the observations on the varieties are random and one or more of the varieties are to be selected.

(b) corresponding to model II, the varieties themselves are drawn by some random process and they would change under replication.

In all cases, a number of observations are made on each of the K varieties. The first attempt towards a suitable formulation for the first category problem is due to Mosteller (33). He gives a non parametric procedure for testing the hypothesis of homogeniety against the slippage alternative that exactly one of the populations has slipped to the right. In other words, one of the varieties is better than the others by a certain amount Δ and the rest are all equal. Briefly, the procedure is as follows. Arrange the observations on all the varieties being considered in order of magnitude and look for the sample with the largest number of observations exceeding all observations on the remaining 'K-1' varieties. If 'r' the number of such observations is greater than 'r_o', a specified number, the result will support the hypothesis of slippage: r_o is chosen to meet the required probability statements.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Following Mosteller (33) Paulson (41) considers the problem of dividing the varieties into superior and inferior groups. The superior group consists of those varieties which fall in the interval $(\bar{X}_m - \lambda \frac{\sigma}{\sqrt{r}} \times \bar{X}_m)$ and the rest of the varieties constitute the inferior group. $\bar{X}_m$ is the average yield of the variety which excells all others, σ is the standard deviation of a single observation, r is the number of replications of each variety and λ is a positive constant. In particular he considers the evaluation of two probabilities H and G; H is the probability of classifying the varieties as unequal when they are actually equal and G is the probability of classifying the varieties equal when they are really different. For specified G, a solution is given for the particular case when all varieties are equal except one and that is better than others by δ^*. Denoting by m_i the mean of the variety V_i it is equivalent to:

$$m = m_1 = \ldots = \underset{i-1}{m} = \underset{i+1}{m} = \ldots = m_k \quad \text{and} \quad m_i = m + \delta^*$$

G may be regarded as the least upper bound of classifying the varieties wrongly when one is better than others by at least δ^*.

In (42) Paulson considers the problem of selecting the best variety out of K in comparison with a control. If we denote V_1 as the standard or the control and its sample mean by $\bar{X}_1$, the procedure suggested is:

if $\bar{X}_m - \bar{X}_1 \geq \lambda \cdot \sigma \cdot \sqrt{\frac{2}{r}}$, select the variety with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the sample mean $\bar{X}_m$ and if $\bar{X}_m - \bar{X}_1 < \lambda \cdot \sigma \cdot \sqrt{\frac{2}{r}}$, select V_1 i.e the control.

$\bar{X}_m$ as before is the sample mean of the variety giving the maximum yield and λ is a positive constant and is determined by the experimenter by the condition that when none of the experimental varieties is better than the control, the probability of V_1 being selected is $1 - \alpha$. Following Paulson's notation in (43) D_0, D_i, $(i = 1, \ldots, K)$ are the decisions corresponding to the situations that all the varieties are equal and that V_i has slipped to the right respectively. He shows that the optimal procedure is:

$$\text{if } \frac{n(\bar{X}_m - \bar{X})}{\sqrt{\sum_{i=1}^{k} \sum_{\alpha=1}^{n} (X_{i\alpha} - \bar{X})^2}} \geq \lambda_\alpha \text{ adopt the decision } D_m$$

$$\text{and if } \frac{n(\bar{X}_m - \bar{X})}{\sqrt{\sum_{i=1}^{k} \sum_{\alpha=1}^{n} (X_{i\alpha} - \bar{X})^2}} < \lambda_\alpha \text{ adopt the decision } D_0$$

λ_α is a constant so determined that the probability of adopting decision D_0 is $1 - \alpha$. The criterion of optimality is that subject to the above restriction the procedure maximizes the probability of D_i when that is the correct decision.

The special problem of selecting a single variety has been discussed from measure theoretic point of view by Bahadur (4). Under certain regularity conditions he shows

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that the procedure of selecting the variety corresponding to the greatest sample mean cannot be improved no matter what the true parameters are.

Instead of the mean, Truax (53) considers the problem of selection based upon variance. He formulates the problem as a slippage problem, i.e., it is a procedure to decide whether the variances are equal or different in the various varieties and if different which of these varieties has slipped to the right. The procedure is to select one of the decisions D_o and D_i, $(i = 1, 2, \ldots K)$. D_o denotes the situation when all the variances are equal and D_i corresponds to the situation when all other variances are equal except that of the ith variety and it is greater than all others. The rule of the procedure is:

$$\text{if} \quad S_m^2 \Big/ \sum_{i=1}^{n} S_i^2 \geqslant L_a \quad \text{select} \quad D_m \text{ but}$$

$$\text{if} \quad S_m^2 \Big/ \sum_{i=1}^{n} S_i^2 < L_a \quad \text{select } D_o \ .$$

m denotes the variety yielding the largest sample variance. L_a is a constant and S^2 has the usual meaning of estimate of variance. The procedure selects one of the $k + 1$ decisions $(D_o , D_1 , , D_k)$ so that (a) when all the variances are equal D_o should be selected with probability $1 - \alpha$. Subject to this condition the procedure is shown to be optimum in the sense that the probability of selecting D_i is maximum when D_i is the correct decision. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

procedure satisfies the conditions of symmetry and invariance under translocation and scale transformations. In his paper of 1960 Truax (54) considers the slippage problem in a more general setting. K populations are characterized by their density functions $f(x_1 - a_1)...$ ..... $f(x_k - a_k)$ and then he restricts to the special case when all a_i are equal except one and that is bigger than the others. Various properties of symmetry, invariance etc. are considered and finally the nature of Baye's solution is considered.

Where as Mosteller, Paulson and Truax formulate the problem as a slippage problem, Bechhoffer (5) gives the ranking formulation. In Bechhoffer's formulation there is no scope for decision D_o, i.e. all the varieties are equal. The varieties are supposed to be unequal and then various types of ranking problems are considered such as selecting the k_n best varieties, k_{n-1} second best and so on. He gives a precise formulation to the problem: Prob $\left[\max (\bar{X}_1, \bar{X}_2, \quad \bar{X}_{n-1}) < \bar{X}_n \right]$, in that he introduces the least favourable configuration of the means and the variances are assumed equal and known. The least favourable configuration is that arrangement of the means in which all the varieties of the inferior group are equal and the same holds for the superior group and the two groups are δ^* apart; δ^* is to be defined by the experimenter. The superior and the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

inferior groups can be of any size. In the extreme case the superior group can consist of only one variety and that amounts to selecting the best variety. Following the same line he gives in (7) a two sample procedure (Stein procedure (46)) for ranking the means when the variances are unknown. From the first sample an estimate of variance S^2 is obtained and on the basis of that a second sample is drawn. The least favourable configuration is introduced to give the minimum probability of correct selection.

Dunnett (12) discusses the problem of comparing several varieties with a control and gives a procedure for making confidence statements about the true values of the k differences $\bar{X}_i - \bar{X}_o$, (i = 1, ... K) $\bar{X}_i$ is the sample mean of variety V_i and $\bar{X}_o$ is the sample mean of the control. The procedure has the property that the probability of all k statements being simultaneously correct is equal to a specified value 'P'. In this formulation the procedure is a special case of the confidence statements for all contrasts considered by Tukey and discussed by Scheffe (60).

Grundy et al (29) give a procedure for adopting or rejecting a new variety in comparison with the standard one. Their procedure assumes that the experiment can be conducted in two stages. On the basis of the first experiment an estimate of the true excess, θ, of the new

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

variety over the standard is obtained and from that the fiducial distribution of 'θ' is also obtained. The decision depends upon the sign of the combined mean of the two stage experiments for θ. If the scale of operations is known the resulting increase in output (gain) can be determined but against that the cost of the experiment has to be balanced. Risk is defined as the cost of the experiment minus expected gain. However, the expected gain depends upon the unknown θ; it is eliminated by integrating over its fiducial distribution as determined from the first experiment. The size of the experiment for the second stage is determined by minimizing the risk as defined above.

Bross (8) deals with the problem of selecting one of the two populations (types of electric bulbs). Three different formulations are given. Firstly, he gives a solution in terms of probabilities for a specified difference δ_o between the means of the two populations. The δ_o is supposed to be defined by the experimenter. The second solution is a minimax solution; that n (sample size) which minimizes the maximum loss is adopted. Loss is a linear function of the difference of the means. Alternative to the minimax approach is suggested the maximization of the gain.

Sommerville (50) treats the problem of optimum sample size for selecting the population having the greatest mean

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

out of k + 1 populations. The optimum size is that which minimizes the maximum loss. The observations on each population are normal with the same variance and loss is a linear function of the difference of the mean of the selected population and the population having the greatest mean.

Seal (61) discusses a class of decision procedures. The basic object of the procedure is that they ensure that the l.u.b. of the probability of not including in the group the variety with the largest mean is 'a'. Subject to this basic consideration the procedure is required to fulfil further conditions of unbiasedness and gradation as defined there. As a special form the procedure comes very close to that of Bechhoffer's (5) method, for example that the best variety will not be included has minimum probability or the worst variety will not be included has the maximum probability.

The problem of selecting from two varieties following a sequential procedure has been discussed by Maurice (38). Test criterion is the probability ratio as used by Wald (55). Loss is a linear function of the absolute difference of the means of the two varieties. Sample size required is obtained by minimax principle. As Maurice (38) shows the sequential procedure is more efficient than that of Grundy et al (29), in the sense that the expected loss is always less in the sequential procedure suggested.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lehman (32) formulates the problem as of dividing the varieties into two groups. Each variety is characterized by a real valued parameter θ. Variety i is taken to be

positive (good) if $\theta_i \geq \theta_o + \Delta$

and negative (bad) if $\theta_i < \theta_o$

The procedure adopted is to select the variety V_i if $T_i > C$; T_i is a suitable statistic based on a sufficient statistic and C is some constant determined by appropriate probability considerations. For measuring the efficiency of the procedure various test criteria are suggested, the following of which are discussed in detail.

(a) the expected number of positives

(b) the expected proportion of true positives i.e. the quantity in (a) divided by the number of positives.

And as a measure of the performance with respect to false positives the criteria are

(i) the expected number of false positives

(ii) expected proportion of false positives.

The procedure when applied to the exponential class of distributions is shown to be optimum in the sense that, subject to the condition that infimum of (a) or (b) is greater than a specified γ, supremum of the quantity in (i) or (ii) is minimum.

Model II problems have been discussed by Anderson (1), Finney (22), Cochran (14), Fairfield Smith (28),

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Davies (15), Armitage et al (3), Birnbaum (9) and Birnbaum and Chapman (10).

Finney (22) considers the problem of selecting a specified fraction of the population. The parent population is assumed to be normal with known variance and mean but the observations are subject to an error (experimental error). The error is supposed to be inversely proportional to the area (resources) devoted to the experiment. The object is to utilize the resources to obtain the maximum gain in the mean value of the selected portion of the population over the parent population. Numerical results are given for two stage selection. From the numerical results tentative conclusions are given to distribute the area, which is supposed to be fixed, between the two stages. The main conclusion for the single stage selection is that unless the experimental error is very small the selection should not be more intense than selecting of .01 fraction of the population and if more intense reduction is required it may be obtained by random culling. For the two stage selection to a first approximation the optimal conditions will be to take $P_1 = P_2 = B^{\frac{1}{2}}$- ($P_1$ is the fraction selected at first stage, P_2 is the fraction selected at second stage and B is the overall fraction selected in two stages).

Cochran (14) again deals with the problem of selecting a fixed fraction of the population. He discusses the case when the selection is effected indirectly i.e. the character

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

desired for has a regression on other traits on which selection is actually based and they are supposed to follow a multivariate normal distribution. An application to plant selection is given which comes very close to what Finney (22) has done. When the selection is effected indirectly the question of constructing the selection index crops up. Generally, a least squares index is used in practice. Cochran shows that the least squares index is not an optimum one. There is always a loss (in gain) by adopting the least squares index but he does not give any alternative. Finally he touches on the problem whether a variate should be discarded or retained for constructing the index.

A variety may be measured in several characters X_1, X_2, .. X_n . A linear function of the X's has been called by Fisher (24) a discriminant function. Fairfield Smith (26) uses the discriminant function for selection. The genotypic (true yielding capacity) value of a variety may be expressed as

$$\psi = \sum_{i=1}^{n} a_i d'_i$$

d'_is are the true values of the characters and a_is are the weights assigned to the characters according to their economic importance. However, d'_is cannot be known exactly but, their estimates can be had, thus the phenotypes will be scored as

$$Y = \sum_{i=1}^{n} b_i X_i$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Y and Ψ are related and the problem Fairfield Smith discusses is to find the b's so that the improvement in genotype is maximum for one stage selection, when a specified fraction is selected.

The gain by selection is the expected gain under the assumption that certain parameters (variances, means etc.) are known. But in practice we can know only estimates of the parameters. Thus with the expected gain there will be attached an error. Nanda (39) following earlier work of Bartlett (58) has given standard error for various forms of the discriminant functions.

To estimate the result of selection the distribution of the parent population is required. Tang (52) gives a method to estimate the density function from the given data. In order to utilize the knowledge about the parent population it is necessary that it should be available before designing the experiment. However, that seems to be impossible and Tang suggests using the data of the past experiments to determine the distribution and use that information to conduct future experiments.

Davies (15) considers the problem of drug screening. Firstly, the difference between drug screening and plant screening is pointed out. In plant selection we are generally restricted to one test per year and practically whole year must pass before they can be tested again. Thus a given proportion of the population can be selected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and tested next year. But in drug screening if there seems to be any active compound it can be tested independently of the others and the time taken is not much. Further it can be assumed that all the compounds have either an activity level zero or 'a' the critical level. A number of criteria of optimum screening are stated and the one which is dealt in detail is P/C; - proportion of active compounds per unit cost. The objective of the whole process is to have the maximum value of P/C with the restriction that the fraction of the population selected is fixed. Armitage & Schniderman (3) have also considered the same problem. However, unlike Davies (15) they do not take the extreme assumption that either the compound has zero activity level or a very high activity level.

Birnbaum (9) considers the same problem from the opposite point of view. Instead of selecting a fraction and evaluating the results of that he is concerned with selection in such a way that the selected population has the desired characters (say means etc.) with the condition that minimum of the population is rejected. $X_1, X_2, \ldots X_k$; $Y_1, \ldots Y_b$ or in Vector notation (XY) are two sets of characters and they all follow a multivariate normal distribution. Selection is to be obtained by truncation for one set of characters - Y's - and its effect on the other set is of real importance. Birnbaum & Chapman (10)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

show that the linear truncation in Y's (i.e. $\Sigma a_i Y_i > t_o$; a's are proper weights) is an optimum procedure for selection. The procedure is said to be optimum if to have certain properties of the Vector X (means etc.) after selection the population rejected is less than that rejected by following any other procedure to obtain the same properties of X.

Finally, Dunnett (19) has discussed the problem of selecting the best of K varieties when the means of the varieties themselves follow normal distribution and that is known a priori. The decision rule depends upon the sample means and the a priori known U's, the means of u's. In case all U's are equal the procedure reduces to the comparison of sample means only.

Statement of the Problem

From the review of the literature it is easy to see that basically the selection problems have two types of formulations. In the first one is concerned with selecting of one or more varieties out of a given number. Various authors have given solutions to this problem under different assumption. Bechhoffer (5) gives the solution assuming the least favourable configuration of the varietal means. Least favourable configuration is that in which k - 1 varieties have equal means and the kth has a mean at least δ^* larger than the others, δ^* is specified by the experimenter. Sommerville (50) arrives at the same configuration but here δ^* is determined by economic costs

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and other constants of his model instead of being specified by the experimenter.

The least favourable configuration serves to cover the cases when no information a priori is available about the means of the varieties. In such cases it is reasonable that the experimenter likes to protect himself against all possible, including the least favourable, configurations. Secondly, such situations as in slippage problems do exist. For all these problems, least favourable configuration is a reasonable assumption.

On the other hand it is reasonable to assume that situations can arise when the experimenter has a priori information about the distribution of the varietal means, e.g. the varietal means themselves follow some distribution - normal, exponential etc.

A typical situation to which this formulation may apply is the selection of the best of k electric bulbs for their brightness and the bulbs are a random selection from the same population. The situation can arise when the a priori distributions have unequal parameters; e.g. if the brightness of the bulbs is under study and the readings have to be taken through a screen, the prior readings on the bulbs may provide an estimate of each bulb separately.

Similar situations can be imagined in agriculture in the selection of new varieties which may be considered as a random selection from a single cohort of newly pro-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

duced strains. Alternatively, already established varieties may be tried in a new region (or country), then from the previous records we have separate estimates of the parameters for all the competing varieties. These situations can occur in all the major agricultural crops but it seems that such a formulation will be particularly suitable for sugarcane in which due to large number of chromosomes and extreme genetic variability every seedling is a potential new variety. A similar example arising in pharmaceutical industry is the selection of antibiotic-producing micro-organisms when these strains are employed in conjunction with a particular raw material, the records on other media give the prior estimates of the antibiotic producing capabilities of the competing strains.

Dunnett (19) has considered one such possibility, that the varietal means follow independently normal distribution with equal and known variance but unequal means. The normal distribution is straightforward to work with, at least much easier than others. However, how far the assumption of normality can be justified in a particular situation can be seen empirically only.

In the first part of the thesis the consequences of non normal situations are evaluated; three non normal distributions, exponential, double exponential and gamma are investigated. In the first instance selection is based on one variate only. In some cases it may be de-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sirable to base selection on more than one character. A formulation of the problem is given when the selection is based on two characters.

The second type corresponds to what is generally named as screening, i.e. selecting only those individuals which exceed in some characters. It is one of the standard techniques for "improving the quality of the population" (of varieties). Beginning with Finney (25) all the authors consider the problem of selecting a specified fraction of the population and almost all of them (except Curnow (63) who has studied some beta distributions) restrict themselves to normal distribution. As Finney (25) has pointed out it will be desirable to introduce in the experiment some already established variety and select only those which are markedly superior to the standard one. In the second part of the thesis consequences of two stage selection, when the selection is in comparison with a standard are evaluated.

In chapter 9 mathematical formulae for a two stage selection, when the selection is based on two characters, are given for the normal distribution. No numerical values have been calculated; in fact the number of variables is so large that it is difficult to tabulate the results in a satisfactory way; however, if the variables are known results can be calculated. These formulae can be extended to more than two characters in a straightforward manner but their numerical evaluation must wait for the general

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

solution of the multiple integrals. For the non normal distributions of the parent population the results of the selection on two characters, if the characters are independently distributed, can be worked out just like the normal and they are briefly indicated. However, it is difficult to define what it means to say that the two characters are individually exponentially (for that sake double exponentially as well) distributed but are not independent.

The usual assumption in the analysis of variance is made that each observation on a variety is the sum of a 'mean' characteristic of the variety and 'e' an error variable with zero mean and known variance. All the e's have the same variance.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PART I

SELECTION

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2

EXPONENTIAL DISTRIBUTION

I. Probability of Correct Selection

On each variety a number of observations are taken, the same number, n, for each of the varieties. The variety with the largest sample mean will be selected. The k unknown population means are denoted u_i and their respective sample means by $\bar{X}_i$ (i = 1, ... k). Each observation on a variety is supposed to be normally distributed with unknown mean (characteristic of the variety) and common known variance σ^2. The statement that V_i is selected implies that

$$\bar{X}_j - \bar{X}_i < 0 \qquad j = 1, \ldots k \neq i$$

$$\text{or} \quad Z_j = \frac{\bar{X}_j - \bar{X}_i - (u_j - u_i)}{\sqrt{(2/n)}\,\sigma} < - \frac{(u_j - u_i)}{\sqrt{(2/n)}\,\sigma}$$

Z_j is a standerdized normal variable with zero mean and unit variance and Z's are correlated with correlation structure

$$\rho\,(Z_j\, Z_k) = \tfrac{1}{2} \qquad j \neq k$$

$$= 1 \qquad j = k$$

Thus p_i, the probability that V_i is selected for given u_i

$$= F_{k-1,\frac{1}{2}} \left[-(u_j - u_i)/\sqrt{(2/n)}\,\sigma; \quad j \neq i \right]$$

$$= F_{k-1,\frac{1}{2}} \left(-y_j/E ; \quad j \neq i \right) \text{, say} \qquad (2\text{-}1)$$

$$E = \sqrt{(2/n)}\,\sigma \quad \text{and} \quad y_j = u_j - u_i$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$F_{k-1,\frac{1}{2}}(Z_1, \ldots Z_{k-1})$ is a cumulative distribution function of 'k-1' variate normal distribution with all correlations equal to $\frac{1}{2}$ and zero means and unit variances.

The probability in (2-1) depends upon u_i which are unknown. Bechhoffer (5) eliminates the unknown u_i by taking them in a least favourable configuration. In this chapter we assume that the means themselves follow independently exponential distributions with parameters $\lambda_1, \ldots \lambda_k$. If all λ_i are equal the population means are a random sample from an exponential distribution. Assuming u's are independent we get the joint distribution of $(u_j - u_i = y_j \quad j \neq i)$ as follows

For negative value of $u_1 - u_i = y_1$, the density function

$$f(u_1 - u_i = y_1 / u_i) = \lambda_1 \exp(-\lambda_1 y_1 - \lambda_1 u_i)$$

defined for $u_i > -y_1$

or $u_i > /y_1/$

and similarly, the density function for general y_j

$$f(u_j - u_i = y_j / u_i) = \lambda_j \exp(-\lambda_j y_j - \lambda_j u_i) \qquad (2\text{-}2)$$

defined for $u_i > -y_j$

For given u_i , y_j are independent and that will give

$$f(y_1, \ldots y_{i-1}, y_{i+1} \ldots y_k / u_i) =$$

$$\prod_{j \neq i} e^{-\sum_{j \neq i} \lambda_j y_j - u_i \sum_{j \neq i} \lambda_j}$$

We know that $f(y_1, .. y_{i-1}, y_{i+1}, .. y_k / u_i) =$

$$\frac{f(y_1, .. y_{i-1}, y_{i+1}, \ldots y_k, u_i)}{f(u_i)}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From that we get

$$f(y_1, \ldots y_{i-1}, y_{i+1}, \ldots y_k, u_i) =$$

$$\prod_{i=1}^{k} \lambda_i \; e^{-\sum_{j \neq i} \lambda_j y_j - u_i \sum_{i=1}^{k} \lambda_i} \qquad (2\text{-}3)$$

The marginal density function of y's can be obtained by integrating out u_i. As noted in (2-2) the density function for each y_j is defined only for values $|y_j| < u_i$, so we can integrate u_i only from max $|y_j|$ or max (minus y_j) to infinity. Thus we have to watch for the configuration of the y's. For k varieties there will be 'k-1' y's and hence (k-1)! configurations of y's are possible. Let us number the different configurations, R_1, R_2, ... $R_{(k-1)!}$ and let R_1 denote the region $|y_1| > |y_2| > \ldots$ $> |y_{k-1}|$. Density function of y's for R_1 will be given by integrating (2-3) for u_i over the range $(-y_1)$ to infinity and it becomes

$$f(y_1, \quad , y_{i-1}, y_{i+1}, \ldots y_k) =$$

$$\int_{-y_1}^{\infty} \prod_{j=1}^{k} e^{-\sum_{j \neq i} \lambda_j y_j - u_i \sum_{j=1}^{k} \lambda_j} \, du_i$$

$$= \frac{\prod_{j=1}^{k} \lambda_j \; e^{-\sum_{j \neq i} \lambda_j y_j + y_1 \sum_{j=1}^{k} \lambda_j}}{\sum_{j=1}^{k} \lambda_j} \qquad (2\text{-}4)$$

defined for $-y_1 > \ldots > -y_{i-1} > -y_{i+1} > \ldots > -y_k$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Thus integrating (2-1) with respect to (2-4) for negative values of y_j will give the probability that u_i is the largest and we have the R_1 configuration of y's and V_i is selected and may be written as

$$R_{1i} = \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \cdots \int_{y_{k-1}}^{\circ} \frac{\prod_{j=1}^{k} \lambda_j \; e^{-\sum_{j\neq i} \lambda_j y_j}}{\sum_{j=1}^{k} \lambda_j} \cdot e^{y_1 \sum_{i=1}^{k} \lambda_i} F_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i)\, dy_k \cdots dy_{i+1}\, dy_{i-1} \cdots dy_1 \tag{2-5}$$

or

$$R_{1i} = \int_{-\infty}^{\circ} \int_{-\infty}^{y_{k-1}} \cdots \int_{-\infty}^{y_2} \frac{\prod_{j=1}^{k} \lambda_j \; e^{-\sum_{j\neq i} \lambda_j y_j}}{\sum_{j=1}^{k} \lambda_j} \cdot e^{y_1 \sum_{i=1}^{k} \lambda_i} F_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i)\, dy_1 \cdots dy_{i-1}\, dy_{i+1} \cdots dy_k \tag{2-6}$$

Summing over all R's will give the probability that u_i is the largest and V_i is selected; and there are $(k-1)!$ configurations (R's) in all. When all λ's are equal (2-5) and (2-6) are greatly simplified and by symmetry all of the R's will give identical values; thereby summation will be replaced by multiplication of R_{1i} by $(k-1)!$, which will give

$$p_i = (k-1)!\; R_{1i} \tag{2-7}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Summing over all i will give the probability of correct selection, i.e. the variety with the greatest mean has been selected and in case λ's are equal all p_i will be equal and that gives the probability of correct selection

$$P = \sum_{i=1}^{k} p_i = k\, p_i$$

$$= k!\ R_{1i} \qquad (2\text{-}8)$$

Each R_i is a 2(k-1) variate integral (not normal) and to give numerical results their solution is required. Dunnett (19) has given numerical results for k equal to 2 only. So for the present we evaluate results for k equal to 2 and compare it with those of Bechhoffer's and Dunnett's. For k equal to 2 (2-8) reduces to

$$\frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\, e^{-\lambda_1 y_1} \int_{-y_1/E}^{\infty} e^{\frac{-t^2}{2}}\, dt\, dy_1 \; +$$

$$\frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\, e^{-\lambda_2 y_2} \int_{-y_2/E}^{\infty} e^{\frac{-t^2}{2}}\, dt\, dy_2 \qquad (2\text{-}9)$$

If λ_1 and λ_2 are equal then (2-9) reduces to

$$\frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} \lambda\ e^{-\lambda y} \int_{-y/E}^{\infty} e^{\frac{-t^2}{2}}\, dt\, dy$$

To indicate the evaluation of (2-4) we solve the first part of the expression,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} e^{-\lambda_1 y_1} \int_{-y_1/E}^\infty e^{\frac{-t^2}{2}} dt\, dy_1$$

Integrating by parts we get

$$\frac{\lambda_2}{\lambda_1+\lambda_2} \left[\tfrac{1}{2} + e^{\lambda_1^2E^2/2} F(-\lambda_1 E)\right] \qquad (2\text{-}10)$$

By symmetry second part of (2-9) is

$$\frac{\lambda_1}{\lambda_1+\lambda_2} \left[\tfrac{1}{2} + e^{\lambda_2^2E^2/2} F(-\lambda E)\right] \qquad (2\text{-}11)$$

If n is equal to zero or E is infinite i.e. no experimentation it can be seen from (2-10) and (2-11) that (2-9) becomes $\frac{1}{2}\left[\lambda_1/(\lambda_1+\lambda_2) + \lambda_2/(\lambda_1+\lambda_2)\right] = \frac{1}{2}$ and that is equivalent to picking any one of the two varieties at random. In case 'n' approaches infinity (2-9) approaches $\left[\lambda_1/(\lambda_1+\lambda_2) + \lambda_2/(\lambda_1+\lambda_2)\right] = 1$, i.e. evidence is so overwhelming that the correct decision is certain. Similar situation can be observed in (2-5) which is the general case. If 'n' is zero the integral has a constant value of $1/k^2(k-1)!$, thus giving the probability of correct selection $1/k$. However, if 'n' is infinite the integral has $1/k(k-1)!$ as its value giving the probability of correct selection unity.

The numerical results can be compared with those of Bechhoffer (5) by taking δ, the difference of the two varieties equal to the expected difference (absolute) between two exponential variates. The expected differ-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ence between two exponentially distributed variables with parameters λ_1 and λ_2 is $(\lambda_1^2+\lambda_2^2)/(\lambda_1+\lambda_2)\lambda_1\lambda_2$ and if λ_1 and λ_2 are equal then the expected difference will be $1/\lambda$.

Take $\lambda_1 = \lambda_2 = \lambda$ and define $T^2 = n\sigma_\delta^2/\sigma^2$ $(=2\sigma_\delta^2/E^2)$. σ_o^2 is the variance of the a priori distribution and σ^2 is the error variance. In the exponential case with parameter λ, σ_o^2 will be $1/\lambda^2$. Thus for the exponential T^2 is equal to $2/\lambda^2E^2$. With these conditions (2-9) reduces to

$$\left[\tfrac{1}{2} + e^{1/T^2}\, I(\sqrt{2}/T)\right]$$

and this depends upon T alone.

Table 1 shows the probability of correct selection for different values of T for normal, exponential, a priori distribution and by Bechhoffer's method by taking δ equal to the expected difference of two exponential variates. The values of T have been taken for which Dunnett (19) has given numerical results for the normal a priori distribution.

Table 1

T	Normal	Exponential	Bechhoffer's method
0	.50	.50	.50
.1584	.55	.51-.52*	.5445
.3249	.60	.5878	.5935
.5095	.65	.6304	.6405
1.00	.75	.7137	.7597
3.078	.90	.8588	.9889
6.314	.95	.9218	more than .9999
31.82	.99	.9827	more than .99999

*the entry is not very accurate. The figure lies between .51 - .52

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The probability of correct selection for the normal *a priori* distribution is consistently greater than that for the exponential case. However, the difference is within 4 percent for every value of T. Thus the exponential gives a useful approximation to the normal case and because of the considerable difference in form between the two distributions the results suggest that the probability of correct selection is not very much dependent on the true form of the *a priori* distribution, at least for k = 2. Probabilities calculated by Bechhoffer's method are greater than those in the exponential case. For values of T less than 1.0 Bechhoffer's method gives slightly lower values than those of the normal, at T equal to 1.0 Bechhoffer's method gives slightly higher value than the normal *a priori* distribution, beyond that the difference increases considerably.

II. Conditional Probability of Correct Selection

In the previous section we considered the probability of correct selection giving no consideration to the particular way the different u's happen to be. That may be requiring too much in some cases. This will be particularly the case when the second best variety does not differ much from the best and the difference does not entail any serious disadvantage if the best one is not selected.

In this section the conditional probability is evaluated,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

i.e. a specified probability of correct selection is required only if the best variety is superior to all others by at least δ. Of course the δ has to be specified by the experimenter by taking into consideration all the relevant circumstances.

From section I (2-1) we have

$$p_i = \text{probability that for given u's } V_i \text{ is selected}$$
$$= F_{k-1,\frac{1}{2}}(-y_{j/E} \; ; \; j \neq i) \qquad (2\text{-}12)$$

and from section I (2-4) we have the joint density function of y's for R_1 configuration of y's

$$\frac{\prod_{j=1}^{k} \lambda_j}{\sum_{j=1}^{k} \lambda_j} e^{-\sum_{j \neq i} \lambda_j y_j} \; e^{y_1 \sum_{j=1}^{k} \lambda_j} \qquad (2\text{-}13)$$

We further require that u_i is at least δ superior to all other u's

$$u_i > \delta + u_j \text{ implies } u_j - u_i < -\delta \qquad j \neq i$$
$$y_j < -\delta$$

Thus we get $\text{prob}(u_i > u_j + \delta) = \Pr(y_j < -\delta)$, for R_1 configuration

$$= \int_{-\infty}^{-\delta} \int_{-y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{k} e^{-\sum_{j \neq i} \lambda_j y_j + y_1 \sum_{i=1}^{k} \lambda_i}}{\sum_{j=1}^{k} \lambda_j} dy_k \ldots dy_{i+1} dy_{i-1} \ldots dy_1 \qquad (2\text{-}14)$$

Thus integrating (2-12) with respect to (2-14) we get the joint probability that u_i is greater than all other u's

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by at least δ and V_i is selected for R_1 region. Dividing this integrated value by (2-14) will give the conditional probability that u_i is greater than all others by at least δ and V_i is selected for the region R_1, and may be written as:

$$R_{1i}^{(\delta)} = \frac{\int_{-\infty}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{K}\lambda_j}{\sum_{j=1}^{K}\lambda_j} e^{-\sum_{j\neq i}\lambda_j y_j} e^{y_i\sum_{i=1}^{K}\lambda_i} F_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i)\,dy_k\cdots dy_{i+1}\,dy_{i-1}\cdots dy_1}{\int_{-\infty}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{K}\lambda_j}{\sum_{j=1}^{K}\lambda_j} \cdot e^{-\sum_{j\neq i}\lambda_j y_j + y_i\sum_{i=1}^{K}\lambda_i} dy_k\cdots dy_{i+1}\cdots dy_{i-1}\cdots dy_1} \tag{2-15}$$

Summing over all R will give the conditional probability that u_i is bigger than all others and V_i is selected. Then summing over all i will give the probability of conditional correct selection and can be written as

$$P = \frac{\sum_{i=1}^{k}\sum_{R}\int_{-\infty}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{K}\lambda_j}{\sum_{j=1}^{K}\lambda_j} e^{-\sum_{j\neq i}\lambda_j y_j + y_i\sum_{i=1}^{K}\lambda_i} F_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i)\,dy_k\cdots dy_{i+1}\,dy_{i-1}\cdots dy_1}{\sum_{i=1}^{k}\sum_{R}\int_{-\infty}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{K}\lambda_j}{\sum_{i=1}^{K}\lambda_j} e^{-\sum_{j\neq i}\lambda_j y_j + y_i\sum_{i=1}^{K}\lambda_i} dy_k\cdots dy_{i+1}\cdots dy_{i-1}\cdots dy_1} \tag{2-16}$$

In case all λ's are equal summation will be replaced by multiplication and the multiplication in the numerator cancels out with the multiplication in the denominator and the conditional probability of correct selection will be equal to (2-15) with all λ's equal.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

II.1.

Again the numerical results can be given for k equal to 2 and compared with normal distribution. For k equal to 2 there is only one R and (2-16) becomes

$$\frac{\frac{1}{\sqrt{2\pi}}\int_{\delta}^{\infty}\lambda_1 e^{-\lambda_1 y_1}\int_{-y_1/E}^{\infty} e^{\frac{-t^2}{2}}\,dt\,dy_1 + \frac{1}{\sqrt{2\pi}}\int_{\delta}^{\infty}\lambda_2 e^{-\lambda_2 y_2}\int_{-y_2/E}^{\infty} e^{\frac{-t^2}{2}}\,dt\,dy_2}{\int_{\delta}^{\infty}\lambda_1 e^{-\lambda_1 y_1}dy_1 + \int_{\delta}^{\infty}\lambda_2 e^{-\lambda_2 y_2}dy_2} \qquad (2\text{-}17)$$

If we take $\lambda_1 = \lambda_2 = \lambda$, it reduces to

$$\frac{1}{\sqrt{2\pi}\ e^{-\lambda\delta}}\int_{+\delta}^{\infty}\lambda e^{-\lambda y}\int_{-y/E}^{\infty} e^{\frac{-t^2}{2}}\,dt\,dy$$

on simplification it gives

$$I(-\delta/E) + I(\delta/E + \lambda E)e^{\lambda^2E^2/2 + \lambda\delta} \qquad (2\text{-}18)$$

If δ is equal to zero (2-18) reduces to $\left[\frac{1}{2} + e^{\lambda^2E^2/2}F(-\lambda E)\right]$, the expression obtained for the probability of correct selection in section I.

In order to compare it with Bechhoffer's method let us fix the value of $(\delta\sqrt{n}/\sigma)$ and that will fix the probability of correct selection when the u's are in the least favourable configuration. From (2-18) it is clear that the probability depends upon λ, δ and $\sqrt{n}/\sigma$. The fixed value of $\delta\sqrt{n}/\sigma$ can be split up into two factors δ

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and $\sqrt{n}/\sigma$ and the probability P can be calculated for different values of $1/\lambda$ i.e. the standard deviation of the exponential distribution.

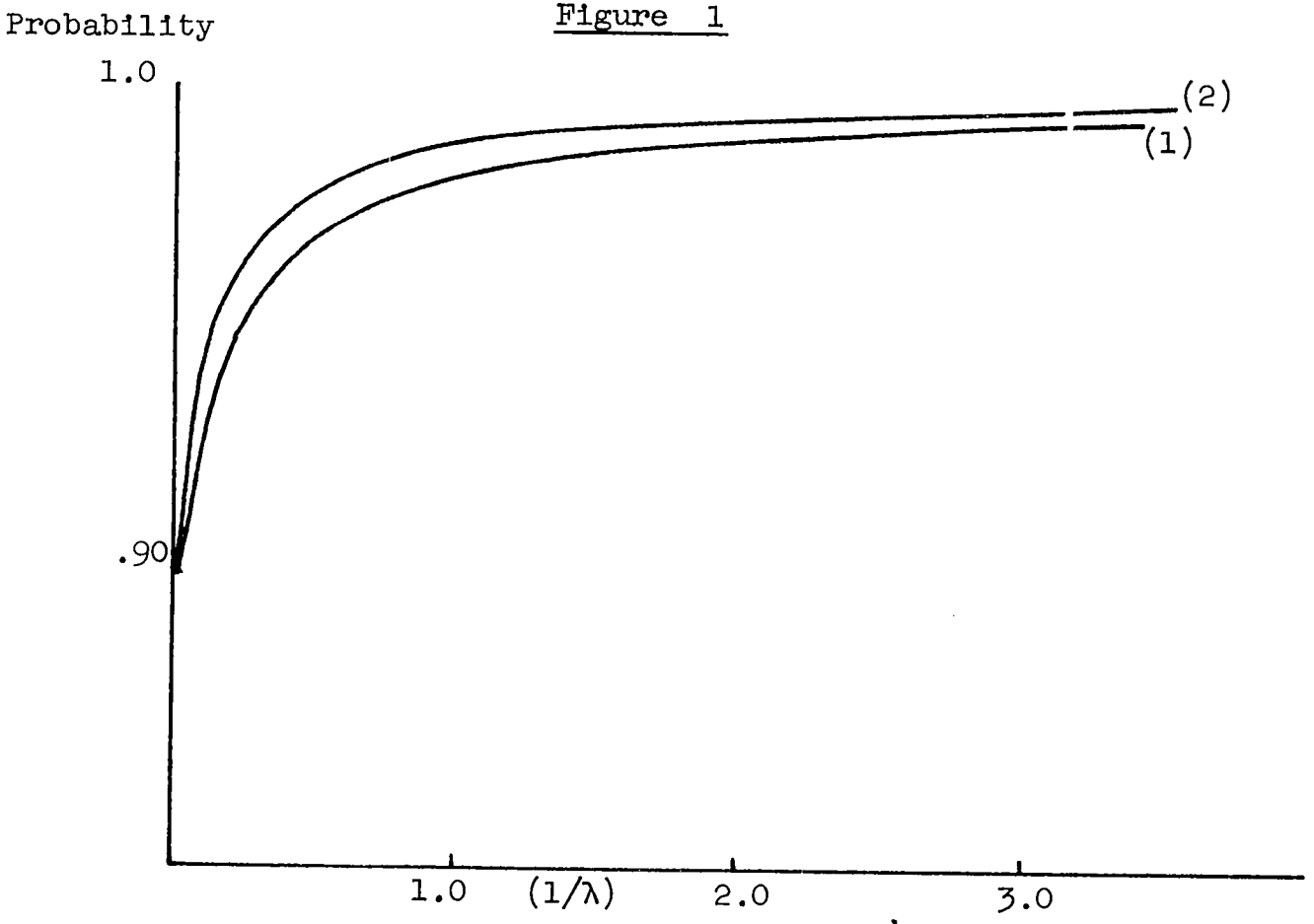
Two values of $\delta\sqrt{n}/\sigma$ have been taken, these correspond to .75 and .90 probability when the means are supposed to be in the least favourable configuration. In each case the value has been split up in two ways; one may be called the narrow ratio of $\delta:\sqrt{n}/\sigma(::1:n)$ and the other wider ones. In both cases the probability is lower for the narrow ratio, though the difference is not large. Figures 1 and 2 show the effect of different splitting. Secondly, as $1/\lambda$ (standard deviation of the exponential a priori distribution) decreases the probability value approaches the one given by Bechhoffer's method. By observing the foot-note * it can be seen from equation (2-18) that when $1/\lambda$ approaches zero its value approaches $I(-\delta/E)$ and that is the same as given by Bechhoffer's method. With the increasing of $1/\lambda$ the probability approaches unity. In order to compare the exponential with the normal the probabilities have been calculated by equation (6') given by Dunnett (19); σ_o, the standard deviation of the normal (a priori) distribution has been equated to $1/\lambda$ and the comparison is shown in figure 3. The normal distribution gives higher values than the exponential though the difference is not very large.

$$* \frac{1}{\sqrt{2\pi}} \int_{\delta/E+\lambda E}^{\infty} e^{-\frac{t^2}{2}} dt \cdot e^{\frac{\lambda^2 E^2}{2} + \lambda\delta} \longrightarrow 0 \quad \text{as } \lambda \longrightarrow \infty$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1



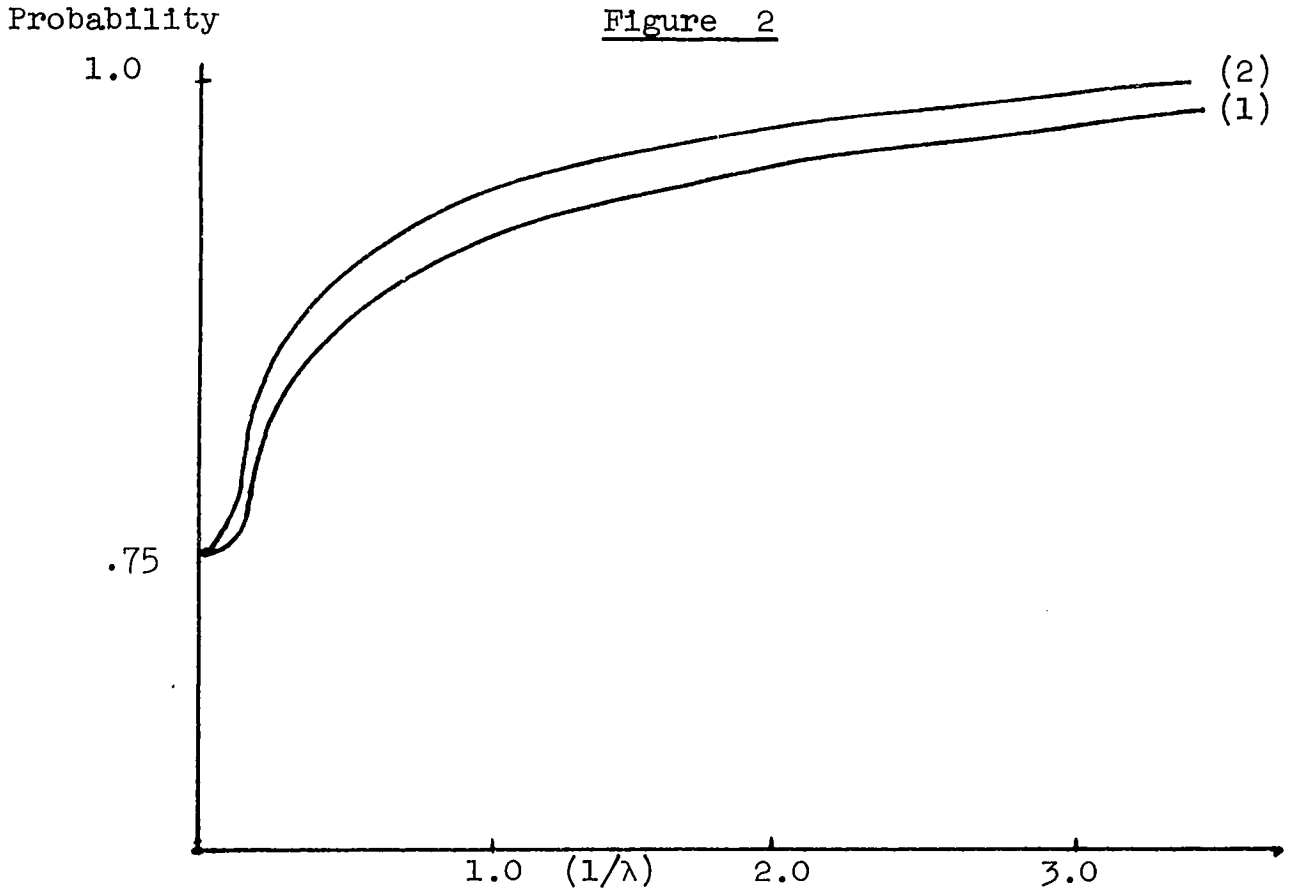
Conditional probability for $\delta\sqrt{(N/\sigma^2)} = 1.8124$

(1) Narrow split; $\delta = .5$ $\sqrt{(N/\sigma^2)} = 3.624$

(2) Wide split; $\delta = .3$ $\sqrt{(N/\sigma^2)} = 6.041$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2



Conditional probability for $\delta\sqrt{(N/\sigma^2)} = .9539$

(1) Narrow split; $\delta = .5$ $\sqrt{(N/\sigma^2)} = 1.9078$

(2) Wide split; $\delta = .3$ $\sqrt{(N/\sigma^2)} = 3.1796$

III.

Another variation of II can be described as follows. Sometimes it is immaterial if the best variety is selected or not as long as the mean of the variety selected is in the neighbourhood of the maximum. So we evaluate the probability that $u_i > u_{max} - \delta$ and u_i is selected. From section I(2-1) we have p_i=Probability that for given u's V_i is selected.

$$= F_{k-1,\frac{1}{2}} (-y_j/E; \quad j \neq i) \quad -- \quad (2\text{-}19)$$

$u_i > u_{max} - \delta$ implies $u_i > u_j - \delta \qquad j \neq i$

$$u_j - u_i = y_j < \delta$$

Integrating (2-19) with respect to joint density function of y_j over the range δ to minus infinity will give the desired probability. But here we run into a difficulty because it is not possible to give a simple joint density function for the y's which is valid in the whole space of the y's. The difficulty can be overcome by determining the density function in different regions separately. The procedure is illustrated for k equal to 3; it can be extended to general k but, beyond $k = 3$ it will become pretty lengthy and tedious. The density function of y's when all are negative has been given in section I (2-4). Here we work out the joint density function of y's when all of them are positive.

$$f(u_j - u_i = y_j/u_i) = \lambda_j e^{-\lambda_j y_j - \lambda u_i}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For given u_i , y_j are independent and thus their density functions can be multiplied. So we get

$$f(u_j - u_i = y_j \; ; \; j \neq i / u_i) = \prod_{j \neq i} \lambda_j \; e^{-\sum_{j \neq i} \lambda_j y_j} \; e^{-u_i \sum_{j \neq i} \lambda_j}$$

and that will give

$$f(y_j \; ; \; j \neq i) = \prod_{j=1}^{k} \lambda_j \cdot e^{-\sum_{j \neq i} \lambda_j y_j} \; e^{-u_i \sum_{i=1}^{k} \lambda_i}$$

Integrating u_i from zero to infinity will give the marginal density of y's and that is

$$f(y_j \; ; \; j \neq i) = \frac{\prod_{j=1}^{k} \lambda_j \; e^{-\sum_{j \neq i} \lambda_j y_j}}{\sum_{i=1}^{k} \lambda_i} \qquad (2\text{-}20)$$

(2-20) may convey the idea that y's are independent, they are not independent. The density function is valid region-wise only because for a given u_i the y's will be ordered as are the u's and there will be $(k-1)!$ regions. When all λ's are equal all the region will be similar.

For k equal to three there will be two y's say y_1 and y_2 and besides the two regions where both are negative or positive there are two other regions; in one of these y_1 will be negative and in the other y_2 will be negative so we need

$$f(y_1 y_2) \quad \text{for} \quad -\infty < y_1 < 0; \quad \infty > y_2 > 0$$

$$\text{and} \quad \text{for} \quad \infty > y_1 > 0; \quad -\infty < y_2 < 0$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the 1st of these regions

$$f(y_1y_2) = \frac{\lambda_1\lambda_2\lambda_3}{\lambda_1+\lambda_2+\lambda_3} e^{2\lambda_1 y_1} e^{-\lambda_2 y_2}$$

and for the second region

$$f(y_1y_2) = \frac{\lambda_1\lambda_2\lambda_3}{\lambda_1+\lambda_2+\lambda_3} e^{-\lambda_1 y_1 + 2\lambda_2 y_2}$$

In the general case in regions other than those in which all y's are +ve or negative we denote the density function as f' with the understanding that for calculating the probabilities f' has to be split up into its proper constituent regions which will depend upon the value of k.

So the probability that the variety V_i has its mean within δ neighbourhood of the maximum and it is selected will be given by the sum of the following three constituents.

(a) Integrated Value of (2-19) with respect to (2-5) over its all regions.

(b) Integrated Value of (2-19) with respect to (2-20) over its all regions. In this case y's are positive but the integration is required for all y's less than δ.

(c) Integrated Value of (2-19) with respect to different constituents of f'.

Taking all λ's equal this may be written as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$p_i=(k-1)!\int_{-\infty}^{\circ}\int_{y_1}^{\circ}\cdots\int_{k-1}^{\circ}\frac{\lambda^{k-1}\,e^{-\lambda\sum_{j\neq i}y_j}}{k}\cdot e^{k\lambda y_1}F_{k-1,\frac{1}{2}}(-y_{j/E};\ j\neq i)\,dy_k\cdots dy_{i+1}\,dy_{i-1}\cdots dy_1$$

$$+(k-1)!\int_{\circ}^{\delta}\int_{\circ}^{y_1}\cdots\int_{\circ}^{k-1}\frac{\lambda^{k-1}\cdot e^{-\sum_{j\neq i}y_j\lambda}}{k}F_{k-1,\frac{1}{2}}(-y_{j/E};\ j\neq i)\,dy_k\cdots dy_{i+1}\,dy_{i-1}\cdots dy_1$$

$$+\int\cdots\int f'\,F_{k-1,\frac{1}{2}}(-y_{j/E};\ )\prod_{j\neq i}^{k}dy_j$$

Summing over all i's will give the probability that any variety has its mean within δ neighbourhood of the maximum and it is selected. It may be noted that for the particular case of k equal to 2, there is no region corresponding to f' and if we take $\lambda_1=\lambda_2=\lambda$ it will give

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\circ}e^{\lambda y}\int_{-\infty}^{-y/E}e^{-\frac{t^2}{2}}\,dt\,dy+\frac{1}{\sqrt{2\pi}}\int_{\circ}^{\delta}\lambda e^{-\lambda y}\int_{-\infty}^{-y/E}e^{-\frac{t^2}{2}}\,dt\,dy \qquad (2\text{-}21)$$

For δ equal to zero it again reduces to the same expression as worked out for probability of correct selection in Section I. And (2-21) simplifies to

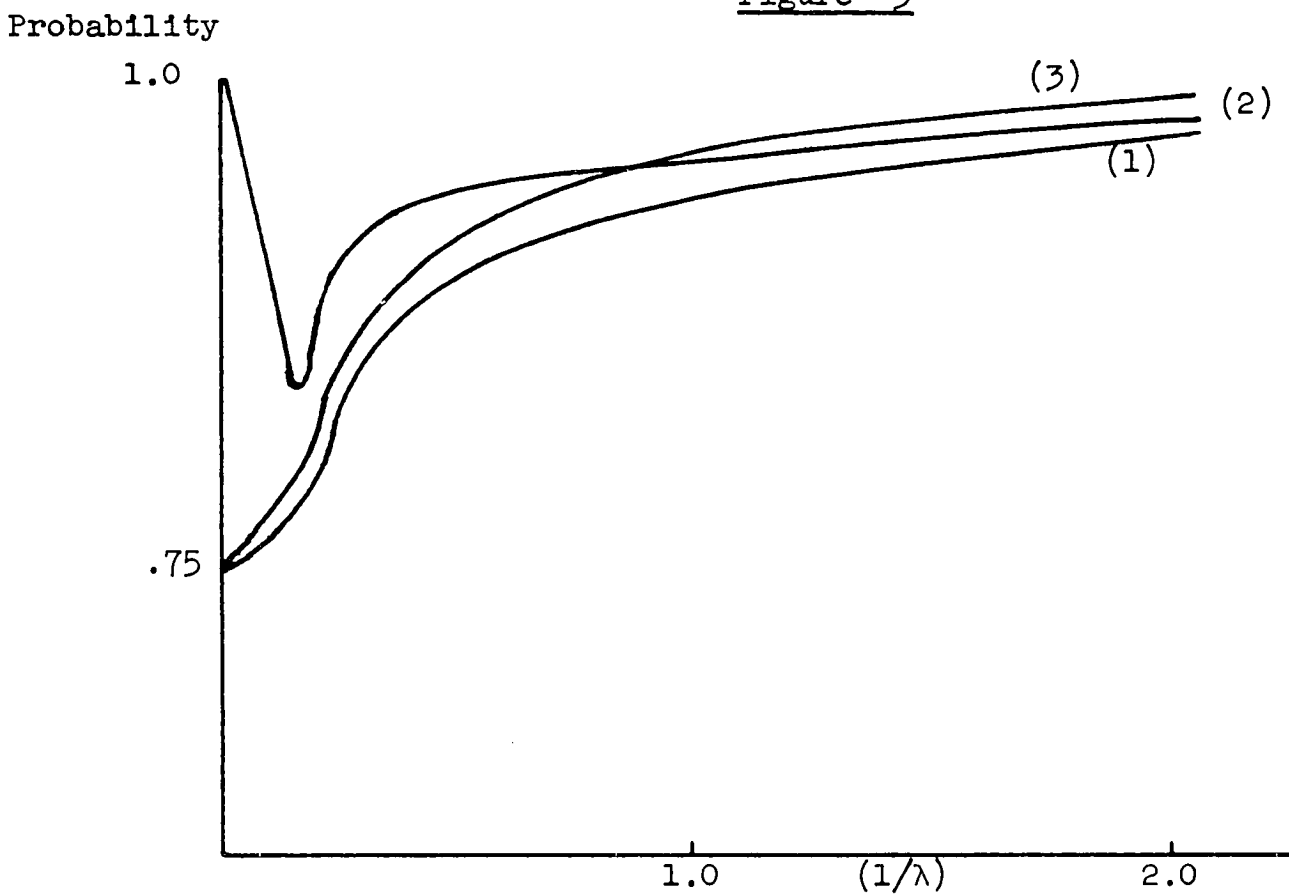
$$1+F(-\lambda E)e^{\frac{\lambda^2E^2}{2}}-F(-\delta/E)e^{-\lambda\delta}-\frac{1}{\sqrt{2\pi}}e^{\frac{\lambda^2E^2}{2}}\int_{\lambda E}^{\delta/E+\lambda E}e^{-\frac{t^2}{2}}\,dt \qquad (2\text{-}22)$$

The above probability can be compared with the conditional probability as defined in the previous section; (2-22) always gives higher values. It can be seen easily

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3



(1) Exponential - Conditional probability of correct selection

(2) Normal - Conditional probability of correct selection

(3) Exponential - Probability that the selected variety is within δ of the best

that as λ approaches infinity the probability defined in (2-22) approaches one and again as λ approaches zero the value approaches one. The curve is shown in figure 3.

IV. An Alternative Approach for Probability of Correct Selection.

The problem of selecting the best variety can be given another form and that is discussed in this section as applied to exponential a priori distribution. Let u_i be the expected response of variety V_i, $(i=1,2,....k)$, where $u_1 > u_2 > .... > u_k$ is an ordered sample from an a priori distribution. For each variety we get a sample of 'n' observations; $\bar{X}_i$ denotes the sample mean, it is normal with mean u_i and variance σ^2/n. We need, for given n, the probability of selecting V_1 - the variety corresponding to u_1 - and that will be equivalent to selecting the best variety or the probability of correct selection.

Probability of selecting the best variety $= \Pr\left[\bar{X}_1 > \max(\bar{X}_2, .. \bar{X}_k)\right]$

$$= P(\bar{X}_1 > \bar{X}_2) - P(\bar{X}_3 > \bar{X}_1 > \bar{X}_2) - P\left[\bar{X}_k > \bar{X}_1 > \max(\bar{X}_2, \bar{X}_{k-1})\right] \qquad (2\text{-}23)$$

$$\Pr(\bar{X}_1 > \bar{X}_2) = F\left[-(u_2 - u_1)/E\right] \qquad (2\text{-}24)$$

$$P(\bar{X}_3 > \bar{X}_1 > \bar{X}_2) = F_{2,-\frac{1}{2}}\left[-(u_2-u_1)/E,\ -(u_1-u_3)/E\right] \qquad (2\text{-}25)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Similarly the other terms can be written down, in general.

$$P\left[\bar{X}_k > \bar{X}_1 > \max(\bar{X}_2, .. \bar{X}_{k-1})\right] = F_{k-1,\frac{1}{2},-\frac{1}{2}}\left[-(u_2-u_1)/E, -(u_3-u_1)/E; -(u_4-u_2)/E\right] \quad (2\text{-}26)$$

where $(\frac{1}{2},-\frac{1}{2})$ denotes that the first k-2 variables are correlated with correlation $\frac{1}{2}$ and the (k-1)th has $-\frac{1}{2}$ correlation with rest of them. The above probability expressions depend upon unknown u's but, they follow exponential distribution and further we know that they are an ordered sample, u_1 is the largest and u_2 the second largest and so on. Joint distribution of max 'u_1' and second max 'u_2' out of a sample of k is given by

$$G(u_1,u_2) = k(k-1)\ dF(u_1)\ dF(u_2)\left[F(u_2)\right]^{k-2} \qquad u_2 < u_1 \quad (2\text{-}27)$$

Similarly joint distribution of max 'u_1' and second max 'u_2' and third max 'u_3' out of a sample of k is given by

$$G(u_1u_2u_3) = k(k-1)(k-2)\ dF(u_1)\ dF(u_2)\ dF(u_3)\left[F(u_3)\right]^{k-3} \quad (2\text{-}28)$$

The other expressions can be written down in a straightforward manner. The probability in (2-24) depends upon u_1 and u_2 and their distribution is given in (2-27). Hence integrating (2-24) with respect to (2-27) will give the probability that the best is selected over the second best. Further we know that u_1 and u_2 originally follow exponential distribution with parameter λ; that will give

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pr (Best is selected over the 2nd best)

$$= k(k-1)\int_0^\infty \lambda e^{-\lambda u_1} \int_0^{u_1} \lambda\, e^{-\lambda u_2}(1-e^{-\lambda u_2})^{k-2} F\left|-(u_2-u_1)/E\right| du_2 du_1 \qquad (2\text{-}29)$$

As it turns out (2-29) is independent of k i.e. for k greater than 2 value of (2-29) is the same as for k equal to 2. The general term in (2-29) to be evaluated will be

$$k(k-1)\int_0^\infty \lambda e^{-\lambda u_1} \int_0^{u_1} \lambda\, e^{-n\lambda u_2} F\left|-(u_2-u_1)/E\right| du_2 du_1 \qquad (2\text{-}30)$$

Little algebra shows that (2-30)

$$= k(k-1)\left|\frac{1}{n+1}\left(\tfrac{1}{2} + e^{\frac{\lambda^2 E^2}{2}}\, I(\lambda E)\right)\right| \qquad (2\text{-}31)$$

For given λE the inner bracket is a constant and so we may write (2-31) as $\dfrac{(k-1)k.P}{n+1.}$ (2-32)

For given k expanding (2-29) and using the result (2-32) it can be written as

$$P.k(k-1)\left|\tfrac{1}{2} - \binom{k-1}{1}\tfrac{1}{3} + \ldots \pm \binom{k-2}{k-2}\tfrac{1}{k}\right| \qquad (2\text{-}33)$$

In (2-33) the terms inside the bracket are equal to $1/k(k-1)$;* thus (2-33) has the value 'P', independent of k.

---

* This can be easily proved by noting the identity

$$(1-\tfrac{1}{t})^{n}\frac{1}{t^3} = \binom{n}{0}\frac{1}{t^3} - \binom{n}{1}\frac{1}{t^4} + \ldots \pm \binom{n}{n}\frac{1}{t^{n+2}}\,.$$

Integrate both sides and substitute 1 for t.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

IV.1.

The other constituents of (2-23) like the first one are also independent of k. We prove it for

$$\text{Prob}\left[\bar{X}_r > \bar{X}_1 > \max(\bar{X}_1, \bar{X}_3, \ldots \bar{X}_{r-1})\right], \text{ where } r < k.$$

From (2-26) we have

$$\Pr\left[\bar{X}_r > \bar{X}_1 > \max(\bar{X}_2, \ldots \bar{X}_{r-1})\right] = F_{k-1,\frac{1}{2},-\frac{1}{2}}\left[-(u_2-u_1)/E, \ldots -\frac{(u_1-u_r)}{E}\right]$$

$$= F_{k-1,\frac{1}{2},-\frac{1}{2}}\left[-\frac{(u_2-u_r)-(u_1-u_r)}{E}, \ldots \frac{-(u_1-u_r)}{E}\right]$$

(2-34)

This depends upon u_i, (i=1, ... r) which are unknown but we know that they are the 1st 'r' ordered observations of a sample of k from an exponential distribution with parameter λ. To prove that (2-34) is independent of k ($k > r$) it will be sufficient to prove that the joint density of

$\left[(u_1-u_r), (u_2-u_r), \ldots (u_{r-1}-u_r)\right]$ is independent of k. To prove this we note that the joint density function of the 1st r ordered observations of a sample of k from an exponential distribution with parameter λ is

$$g(u_1, \ldots u_r) = (1-e^{-\lambda u_r})^{k-r} \cdot \lambda^r \cdot e^{-\lambda(u_1+u_2+\ldots+u_r)} \frac{k!}{(k-r)!}$$

$$0 < u_r < u_{r-1} < \ldots < u_1$$

And the density function of u_r

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$g(u_r) = \lambda(1-e^{-\lambda u_r})^{k-r}\, e^{-r\lambda u_r}\, \frac{k!}{(k-r)!\ (r-1)!}$$

Further we have the conditional density function of $(u_1, \ldots, u_{r-1})$ for given u_r

$$g(u_1,\ldots u_{r-1}/u_r) = \frac{g(u_1,\ldots u_r)}{g(u_r)} = \lambda^{r-1} e^{-\lambda(u_1+u_2+\ldots+u_{r-1})}\, e^{\lambda(r-1)u_r}$$

(2-35)

Let $y_i = u_i - u_r$, for $(i = 1,\ldots,r-1)$ y_i is always greater than zero. From (2-35) we get

$$f(y_i\ ;\ i=1,\ldots r-1/u_r) = \lambda^{r-1}\, e^{-\lambda\left| y_1+y_2+\ldots+y_{r-1}+\lambda(r-1)u_r \right|} e^{+\lambda u_r(r-1)} \times (r-1)!$$

(2-36)

Marginal density function of y_i's will be given by multiplying (2-36) by the density function of u_r as given earlier and integrating out u_r and it reduces to

$$f(y_i\ ,\ i=1,\ldots r-1) = \lambda^{r-1}\, e^{-\lambda(y_1+y_2+\ldots+y_{r-1})}(r-1)!$$

which is independent of k, it really depends upon r only.

To evaluate other than the first constituents of (2-23) we have to evaluate multiple integrals. However the upper bounds for all these can be obtained, that will involve univariate integrals only.

$$\Pr(\bar{X}_r > \bar{X}_1) > \Pr\left[\bar{X}_r > \bar{X}_1 > \max(\bar{X}_2, \ldots \bar{X}_{r-1})\right]$$

$$\Pr(\bar{X}_r > \bar{X}_1) = F\left[-(u_1-u_r)/E\right] = F(-y/E). \qquad (2\text{-}37)$$

u_1 and u_r are the largest and the rth largest ordered observations from a sample of k. Density function of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$y = u_1 - u_r$ $(y > 0)$ is independent of k and can be worked out in a straightforward manner and it is

$$(r-1)\,\lambda\,(1-e^{-\lambda y})^{r-2}\,e^{-\lambda y} \qquad \infty > y > 0 \tag{2-38}$$

Integrating (2-37) with respect to (2-38) will give the probability that $\bar{X}_r > \bar{X}_1$

$$= (r-1)\int_0^{\infty} (1-e^{-\lambda y})^{r-2}\, e^{-\lambda y}\, F(-\tfrac{y}{E})\, dy \tag{2-39}$$

IV.2.

From the observation that the various constituents of (2-23) are independent of k we obtain a very interesting result and that is derived below. As given earlier

$$\text{Prob(correct selection)} = P(\bar{X}_1 > \bar{X}_2) - P(\bar{X}_3 > \bar{X}_1 > \bar{X}_2) - \ldots - P\left[\bar{X}_k > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{k-1})\right]$$
$$= 1 - P(\bar{X}_2 > \bar{X}_1) - \ldots - P\left[\bar{X}_k > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{k-1})\right] \tag{2-40}$$

If we take the sample size 'n' equal to zero i.e. no experimentation we get

$$\text{Prob(correct selection)} = 1 - (\tfrac{1}{2} + \tfrac{1}{3!} + \tfrac{2!}{4!} + \ldots + \tfrac{(k-2)!}{k!}) = \tfrac{1}{k}.$$

This can be easily checked and the terms within the brackets are obtained by combinatorial method e.g. $\Pr\left[\bar{X}_r > \bar{X}_1 > (\bar{X}_2, \ldots \bar{X}_{r-1})\right]$ is given by the permutations in which we have the particular configuration divided by all possible permutations of r objects. The effect of increasing n is an increase in the probability of correct selection and that is obtained by a decrease in the terms inside the brackets. For a given n we may write

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$P_{k_1}(n)$ = Probability correct selection out of k_1 varieties

$$= 1-\left(\frac{1}{2} + \frac{1}{3!} + \ldots + \frac{(k_1-2)!}{k_1!} - d(k_1)\right) = \frac{1}{k_1} + d(k_1),$$

Where $d(k_1)$ is the decrease in the terms within the bracket. Following the same argument we write

$$P_{k_2}(n)=1 - \left(\frac{1}{2} + \frac{1}{3!} + \ldots \frac{(k_2-2)!}{k_2!} - d(k_2)\right) = \frac{1}{k_2} + d(k_2)$$

where $d(k_2)$ is defined analogous to $d(k_1)$. From the fact that the constituents of (2-40) are independent of k, we get that $d(k_2)$ will be always greater than $d(k_1)$. Thus replacing $d(k_2)$ by its lower bound, we get

$$P_{k_2}(n) > P_{k_1}(n) - \frac{1}{k_1} + \frac{1}{k_2}$$

$$= P_{k_1}(n) + \frac{k_1-k_2}{k_1k_2} \qquad (2\text{-}41)$$

This holds for all $k_2 > k_1$. The second term in (2-41) is always negative and its numerical value is less than $\frac{1}{k_1}$. So if for k of the order of 5 tables are available, those will be sufficient for all k, for most of the practical purposes. The probability given in (2-41) is the lower limit. The approximation can be further improved as below.

Let us suppose we have the table for k_1

Then $P_{k_1}(n) = \frac{1}{k_1} + d(k_1)$.

$$d(k_1)=\left[.5-Pr(\bar{X}_2 > \bar{X}_1)\right]+\ldots+\left[\frac{(k_1-2)!}{k_1!} - P(\bar{X}_{k_1} > \bar{X}_1 > (\bar{X}_2..\bar{X}_{k-1})\right]$$

Likewise

$$d(k_2)= \left[.5-Pr(\bar{X}_2 > \bar{X}_1)\right]+\ldots+\left[\frac{(k_1-1)!}{(k_1+1)!} - P(\bar{X}_{k+1} > \bar{X}_1 > (\bar{X}_2,..\bar{X}_{k+1})\right]+..$$

$$+\left[\frac{(k_2-2)!}{k_2!} - P(\bar{X}_{k_2} > \bar{X}_1 > (\bar{X}_2,..\bar{X}_{k_2-1})\right]$$

$$D(k_2) = d(k_2) - d(k_1) > 0$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$= \left[\frac{(k_1-1)!}{(k_1+1)!} - P(\bar{X}_{k_1+1} > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{k_1+1}) \right] + \ldots + \left[\frac{(k_2-2)!}{k_2!} - P(\bar{X}_{k_2} > \bar{X}_1 > (\bar{X}_2 - \bar{X}_{k_2-1}) \right]$$

To evaluate $D(k_2)$ we need the value of multiple integrals but we can get its lower limit and that will be given by having the upper limit of the terms like $P\left[\bar{X}_r > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{r-1})\right]$ etc., that has already been explained in section IV.1. If k_1 is even moderate, say 5 or 6, and $P(k_1)$ is fairly high, say of the order of 80%, the lower bound of $D(k_2)$ will be fairly close to its true value.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3

Gamma and Double Exponential Distributions

In this chapter we investigate the gamma and the double exponential distributions. In the first part three probabilities; probability of correct selection, conditional probability of correct selection and the probability that the selected variety is within δ of the maximum for the gamma are considered. This will include χ^2 distribution as well because χ^2 is a special case of gamma with λ equal to $\frac{1}{2}$. Second part deals with double exponential distribution.

3.1 Gamma

I. Probability of Correct Selection

Assuming the u's follow independently gamma distribution with given r (r can be different for every u_i but in sequel it will be assumed to be the same for all u's) and parameters $\lambda_1, \lambda_2 \ldots \lambda_k$, we can get the joint density function of $u_j - u_i = y_j$. The joint density function worked out for y_j is valid regionwise only. In different regions the joint density function will be different. For this problem we need the joint density function of y's when all of them are less than zero.

Density function of $u_j - u_i = y_j$, for negative value of y_j is given as

$$f(u_j - u_i = y_j \mid u_i) = \frac{\lambda_j^r}{\Gamma(r)} (u_i + y_j)^{r-1} e^{-\lambda_j y_j - \lambda_j u_i} \qquad (3\text{-}1)$$

defined for $u_i > -y_j$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For given u_i, y_j are independent and that will give

$$f(y_j;\ j\neq i \mid u_i)=\frac{\prod_{j\neq i}^{k}\lambda_j^{r}\ \prod_{j\neq i}^{k}}{\left[\Gamma(r)\right]^{k-1}}(u_i+y_j)^{r-1}e^{-\sum_{j\neq i}^{k}\lambda_j y_j-u_i\sum_{j\neq i}^{k}\lambda_j}$$

And hence

$$f(y_j,u_i\neq j\neq i)=\frac{\prod_{j=1}^{k}\lambda_j^{r}\ \prod_{j\neq i}^{k}}{\left[\Gamma(r)\right]^{k-1}}(u_i+y_j)^{r-1}\,u_i^{r-1}\,e^{-\sum_{j\neq i}^{k}\lambda_j y_j-u_i\sum_{j=1}^{k}\lambda_j} \qquad (3\text{-}2)$$

The marginal density function of y's can be obtained by integrating out u_i over the range max $|y_j|$ to infinity. In different regions, for negative y's there will be (k-1)! of them, the integration will be over different ranges. Thus for a particular region, say R_1, when y_1 is the largest, the density function will be,

$$f(y_j;\ j\neq i)=\int_{-y_1}^{\infty}\frac{\prod_{j=1}^{k}\lambda_j^{r}}{\left[\Gamma(r)\right]^{k}}\prod_{j\neq i}^{k}(u_i+y_j)^{r-1}\,u_i^{r-1}\,e^{-\sum_{j\neq i}^{k}\lambda_j y_j-u_i\sum_{j=1}^{k}\lambda_j}\,du_i$$

$$= f_{1i} \qquad \text{say} \qquad (3\text{-}3)$$

For given u_i the probability that V_i is selected is given by (2-1) of section I chapter 2. Integrating (2-1) with respect to (3-3) over the negative values of y's will give the joint probability that u_i is greater than all other u's and V_i is selected for the region R_1, and it may be written as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$R_{1i} = \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \cdots \int_{y_{k-1}}^{\circ} f_{1i}\, F_{k-1,\frac{1}{2}}\,(-y_j/E\ ;\ j\neq i)\,dy_k \cdots dy_{k+1}\, dy_{i-1} \cdots dy_1 \qquad (3\text{-}4)$$

Summing over all R and i will give the probability of correct selection.

For given r and k the expression in (3-3) can be worked out. But to evaluate (3-4) will require the values of multiple integrals and their combined moments, the order of which will depend upon k and r. For the particular case of k equal to 2 the numerical results can be calculated and compared with other distributions.

For k equal to 2 and taking $\lambda_1=\lambda_2=\lambda$ (3-3) reduces to

$$\int_{-y}^{\infty} \frac{\lambda^{2r}}{\left[\Gamma(r)\right]^2}\,(u+y)^{r-1}\,u^{r-1}\,e^{-\lambda y-2\lambda u}\,du$$

$$= \frac{\lambda^{2r}\cdot e^{\lambda y}}{\left[\Gamma(r)\right]^2}\left[\frac{(-y)^t}{2\lambda} + \frac{t\,(-y)^{t-1}}{(t-1)!\,(2\lambda)^2} + \frac{t!\,(-y)^{t-2}}{(t-2)!\,(2\lambda)^3} + \cdots + \frac{t!}{(2\lambda)^{t+1}}\right]$$

$$+\binom{r-1}{1} y\left[\frac{(-y)^{t-1}}{2\lambda} + \frac{(t-1)!\,(-y)^{t-2}}{(t-2)!\,(2\lambda)^2} + \cdots + \frac{(t-1)!}{(2\lambda)^t}\right]$$

$$+\cdots+\binom{r-1}{r-1}\, y^{r-1}\left[\frac{(-y)^{r-1}}{2\lambda} + \frac{(-y)^{r-2}}{(2\lambda)^2}\,\frac{(r-1)!}{(r-2)!} + \cdots + \frac{(r-1)!}{(2\lambda)^r}\right] \qquad (3\text{-}5)$$

$$= \frac{\lambda^{2r}}{\left[\Gamma(r)\right]^2}\cdot e^{\lambda y}\,(f) \qquad \text{say}$$

$$t = 2r - 2$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

If we define $\binom{N}{n} = 0$ for $n > N$ and adopt the convention that $(-y)^n$ for n less than zero is zero (rather undefined) it is clear that for the particular case of r equal to one the first term within brackets is $\frac{1}{2}\lambda$ and others reduce to zero and (3-5) reduces to $\frac{\lambda}{2} e^{\lambda y}$ and that is what we get for the exponential in section I of chapter 2.

For k equal to 2 the probability of correct selection will include two terms only and when $\lambda_1 = \lambda_2$ both the terms will have identical values thus it is sufficient to evaluate one term only, P

$$P = \frac{\lambda^{2r}}{2\lambda \left| \Gamma(r) \right|^2} \int_{-\infty}^{\circ} e^{\lambda y}(f)\ F(-y/E)\ dy \qquad (3\text{-}6)$$

(3-6) is the sum of several terms, the number of which depends upon r. The general term apart from constants can be written as

$$T_n = \int_{-\infty}^{\circ} e^{\lambda y}\ y^n F\ (-y/E)\ dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\circ} y^n e^{\lambda y} \int_{-\infty}^{-y/E} e^{-\frac{t^2}{2}}\ dt \cdot dy \qquad (3\text{-}7)$$

Integrating (3-7) by parts gives a recursive relation of the form

$$T_n = \frac{-n\ T_{(n-1)}}{\lambda} + \frac{e^{\lambda^2 E^2}}{\lambda}\ t_n$$

where $t_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\circ} y^n\ e^{-(y-\lambda E^2)^2/2E^2}\ dy$

Similarly the recursive relation for t_n can be obtained by integrating by parts and it is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$t_n = (n-1)E^2 t_{n-2} + \lambda E^2 t_{(n-1)}.$$

In order to evaluate T_n apart from the recursive relations we need T_o, t_o and t_1 as basic quantities to start the process

$$T_o = \frac{1}{2\lambda} + \frac{1}{\lambda} e^{\frac{\lambda^2 E^2}{2}} F(-\lambda E)$$

$$t_o = F(-\lambda E)$$

and $$t_1 = -\frac{E - e^{-\lambda^2 E^2/2}}{\sqrt{2\pi}} + \lambda E^2 F(-\lambda E)$$

The following two tables show some features of the values obtained for the gamma distribution.

Table 2

Probability of correct selection for k=2

r	gamma	Normal with matching variance	Least favourable
1	.6683	.6960	.6888
2	.7306	.7501	.7732
3	.7676	.7820	.8258
4	.7928	.8039	.8933
8	.8466	.8522	.9856

These values are for $\lambda = .5$ and $E = 4.0$

The values for least favourable configuration have been obtained by taking δ equal to the expected difference of the two gamma variables. From table 2 we can observe that the probability of correct selection increases with increasing

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'r' but, it increases faster for the least favourable method; that should be expected because with increasing 'r' the expected difference increases rapidly. In chapter 2 section I we conjectured that, for k=2, the probability of correct selection is not very much dependent upon the a priori distribution. From table 2 we can easily see that the probability for the normal is greater than that for the gamma but difference is not very large and that lends support to the forementioned conjecture.

Table 3 has been constructed in such a way that λ and r change but the mean (r/λ) remains constant.

Table 3

λ	r	mean	E=1.0	E=.5	E=.25
.25	1	4	.7616	.8495	.9760
.5	2	4	.7306	.8302	.9625
1.0	4	4	.7767	.7456	.9688
.25	2	8		.9055	
.5	4	8		.9335	
1.0	8	8		.9341	

From this we observe that the probability increases for every λ with decreasing of E, i.e. as the error variance decreases probability increases. As for the increase in 'r' under the above restriction it appears that eventually with increasing 'r' probability starts increasing.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

II. Conditional probability of correct selection.

As in section II of chapter 2 we require the probability of correct selection only if the best variety is superior to others by δ. Going through similar arguments as in section II of chapter 2 we have

$$\text{Prob}\,(u_i > u_j + \delta) = \Pr\,(y_j < -\delta)$$

$$= \int_{-\infty}^{-\delta} \int_{y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} f_{1i}\; dy_k \ldots dy_{k+1}, dy_{i-1} \ldots dy_1 \qquad (3\text{-}8)$$

for the region R_1, where f_{1i} is defined in (3-3). Integrating (3-1) with respect to (3-8) and dividing that by (3-8) will give the joint conditional probability that u_i is greater than all other u's by δ and V_i is selected for the region R_1. It may be written as

$$R_{1i} = \frac{\int_{-\infty}^{-\delta} \int_{y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} f_{1i}\; F_{k-1,\frac{1}{2}}\,(-y_{j/E}\,;\; j \neq i)\; dy_k \ldots dy_1}{\int_{-\infty}^{-\delta} \int_{y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} f_{1i}\; dy_k \ldots dy_1} \qquad (3\text{-}9)$$

And the conditional probability of correct selection will be given by summing over all R and i, it is

$$P = \frac{\sum_{i=1}^{k} \sum_{R} \int_{-\infty}^{-\delta} \int_{y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} f_{1i}\; F_{k-1,\frac{1}{2}}(-y_j/E\;;\; j \neq i)\, dy_k \ldots dy_1}{\sum_{i=1}^{k} \sum_{R} \int_{-\infty}^{-\delta} \int_{y_1}^{-\delta} \cdots \int_{y_{k-1}}^{-\delta} f_{1i}\; dy_k \ldots dy_1} \qquad (3\text{-}10)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Again the numerical results can be given for k equal to 2 only. For this particular case the numerator as well as the denominator will have two terms and as $\lambda_1 = \lambda_2$ the two terms in the numerator will be equal and same holds for the denominator, thus the net result depends upon the single term only.

$$P = \frac{\frac{\lambda^{2r}}{|\Gamma(r)|^2} \int_{-\infty}^{-\delta} e^{\lambda y} (f) \; F(-y/E)dy}{\frac{\lambda^{2r}}{|\Gamma(r)|^2} \int_{-\infty}^{-\delta} e^{\lambda y}(f) \; dy} \qquad (3\text{-}11)$$

The general term in the numerator apart from constants can be written as

$$T_n = \int_{-\infty}^{-\delta} e^{\lambda y} \cdot y^n \cdot F(-y/E)dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\delta} e^{\lambda y} \; y^n \int_{-\infty}^{-y/E} e^{-\frac{t^2}{2}} dt \cdot dy$$

and the corresponding recursive relations are

$$T_n = F(\delta/E) \; e^{-\lambda\delta} \; \frac{(-\delta)^n}{\lambda} - \frac{n \; T_{n-1}}{\lambda} + \frac{e^{\frac{\lambda^2 E^2}{2}}}{\lambda} - t_n$$

where $t_n = \frac{1}{\sqrt{2\pi}} \; \frac{1}{E} \int_{-\infty}^{-\delta} y^n \; e^{-\frac{(y-\lambda E^2)^2}{2E^2}} \; dy$

and for t_n the recursive relation is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$t_n = \frac{-E}{\sqrt{2\pi}} \; e^{-\frac{(\delta+\lambda E^2)^2}{2E^2}} \; (-\delta)^{n-1} + E^2(n-1)\,t_{n-2} + \lambda E^2 t_{n-1}$$

The basic quantities needed for the calculation of the numerator are T_o, t_o and t_1.

$$T_o = \frac{e^{-\delta\lambda}}{\lambda} \; F(\delta/E) + \frac{e^{\lambda^2 E^2/2}}{\lambda} \; F\left[-\frac{(\delta+\lambda E^2)}{E}\right]$$

$$t_1 = \frac{-E}{\sqrt{2\pi}} \; e^{-\frac{(\delta+\lambda E^2)^2}{2E^2}} + \lambda E^2 \; F\left[-\frac{(\delta+\lambda E^2)}{E}\right]$$

and $$t_o = F\left[-\frac{(\delta+\lambda E^2)}{E}\right]$$

For the denominator general term (apart from constants) is

$$D_n = \int_{-\infty}^{-\delta} e^{\lambda y} \; y^n\text{-}dy = (-\delta)^n \; \frac{e^{-\lambda\delta}}{\lambda} - \frac{n}{\lambda} \cdot D_{n-1}$$

$$D_o = \frac{e^{-\lambda\delta}}{\lambda}$$

Tables 4 and 5 show some features of conditional probability for the gamma distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4

Conditional probability of correct selection for $\lambda=1$, $\delta=.3$

r	Probability	
	E = 2.0	E = 1.0
1	.7171	.7560
2	.7637	.8096
3	.7990	.8380
4	.8179	.8517
5	.8433	....

The conditional probability increases with an increase in r. In chapter 2 for the exponential distribution we noted that as $1/\lambda$ increases conditional probability approaches unity. Similar behaviour can be observed from the above table. With smaller E the conditional probability is greater.

Table 5 has been constructed in such a way that λ and r change but mean remains the same. Then we notice that with smaller λ probability is higher than the corresponding mean with greater λ. This fact can be utilized to determine the upperbound of the conditional probability for gamma distribution by means of exponential with λ reduced accordingly.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5

For $\delta = .3$ $E = 2.0$

λ	r	Probability
.2	1	.8990
1.0	5	.8433
.25	1	.8745
1.0	4	.8179
.5	1	.8070
1.0	2	.7637

The gamma distribution with r = 1 is equivalent to exponential; thus table 5 gives a comparison of the two distributions:

III. Probability that the selected variety is within δ neighbourhood of the best

In certain cases it may be immaterial as long as the mean of the selected variety is in the neighbourhood of the best one. In this section we deal with such a situation for the gamma distribution. Probability that for given u's V_i is selected is given in (2-1).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$u_i > u_{max} - \delta \Rightarrow u_i \geq u_j - \delta$$

$$u_j - u_i = y_j < \delta \quad \text{for all } j \neq i$$

The joint density function of y's for all y's less than zero is given in (3-3). For the evaluation of the probability, the density function of y's when they are all positive is also needed.

Density function of $u_j - u_i = y_j$, for positive y_j is given as

$$f(u_j - u_i = y_j \neq u_i) = \lambda_j^r \ (u_i + y_j)^{r-1} \ e^{-\lambda_j y_j - \lambda_j u_i}$$

$$\infty > u_i > 0$$

For given u_i, y_j are independent and moreover y_j are ordered, the ordering will depend upon the region, when λ is the same all the regions will be similar and there will be $(k-1)!$ of them.

$$f(y_j ; \ j \neq i / u_i) = \frac{\prod_{j \neq i} \lambda_j^r \ \prod_{j \neq i}^{k}}{\left[\Gamma(r)\right]^{k-1}} (u_i + y_j)^{r-1} \ e^{-\sum_{j \neq i}^{k} \lambda_j y_j - u_i \sum_{j \neq i}^{k} \lambda_j} \qquad (3\text{-}12)$$

Multiplying (3-12) by the density function of u_i and integrating out u_i will give the density function of y_j

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$f(y_j\ ;\ j\neq i)=\int_0^\infty \frac{\prod_{j\neq i}^{k}\lambda_j^r}{\left|\Gamma(r)\right|^k}\prod_{j\neq i}^{k}(u_i+y_j)^{r-1}\,u_i^{r-1}\,e^{\sum_{j\neq i}^{k}\lambda_j y_j-u_i\sum_{j=1}^{k}\lambda_j}\,du_i \qquad (3\text{-}13)$$

$$= f_i'$$

For k=2 and taking $\lambda_1=\lambda_2=\lambda$ (3-13) reduces to

$$f(y)=\frac{\lambda^{2r}}{2\lambda\left|\Gamma(r)\right|^2}\,e^{-\lambda y}\left[\frac{\binom{r-1}{0}(2r-2)!}{(2\lambda)^{2r-2}}+\frac{\binom{r-1}{1}(2r-3)!}{(2\lambda)^{2r-3}}+\ldots+\frac{\binom{r-1}{r-1}y^{r-1}(r-1)!}{(2\lambda)^{r-1}}\right] \qquad (3\text{-}14)$$

$$= f', \quad \text{say}$$

If we define $\binom{N}{n}=0$ for $n > N$ it is clear that for the particular case of r equal to one the first term within brackets is unity and others reduce to zero and (3-14) reduce to $\frac{\lambda\, e^{-\lambda y}}{2}$, and that is the same as we got for exponential.

For other regions in which some of the y's are positive and others negative the integration for u_i will be from $|y_{max}|$ of those y's which are negative. Let us call the resulting density function by f_i''. Therefore the joint probability that u_i is within δ of the max u_i or V_i is selected is given by integrating (2-1) with respect to (3-3) over the range zero to minus infinity, plus the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

integrated value of (2-1) with respect to (3-13) over the range zero to δ, plus the integrated value of (2-1) with respect to f''_i the last case the range of integration will depend upon the particular region. The probability P_i, when all λ's are equal, may be written as

$$Pi = (k-1)! \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \cdots \int_{y_{k-1}}^{\circ} f_{1i} \, F_{k-1,\frac{1}{2}}(-y_{j/E}; j \neq i) dy_k \cdots dy_1$$

$$+ (k-1)! \int_{\circ}^{\delta} \int_{\circ}^{y_1} \cdots \int_{\circ}^{y_{k-1}} f''_i \, F_{k-1,\frac{1}{2}}(-y_{j/E}; j \neq i) dy_k \cdots dy_1$$

$$+ \int f''_i \, F_{k-1,\frac{1}{2}}(\quad) \pi dy_j \qquad (3\text{-}15)$$

However for the particular case of k equal to 2 there is no region corresponding to density function f" and on account of the assumption of $\lambda_1 = \lambda_2 = \lambda$ the two terms will be equal. Thus the probability of correct selection in the sense that $u_i > u_{max} - \delta$

$$= 2 \int_{-\infty}^{\circ} f_i F(-y/E) dy + 2 \int_{\circ}^{\delta} f'_i f(-y/E) dy$$

$$= \frac{2 \lambda^{2r}}{\left[\Gamma(r)\right]^2} \int_{-\infty}^{\circ} e^{\lambda y} (f) F(-y/E) dy + 2 \int_{\circ}^{\delta} f' F(-y/E) dy$$

(3-15-a)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where f and f' are defined in (3-5) and (3-14) respectively. The first part in (3-15) is the same as in (3-6). In the second part the general term apart from constants can be denoted as

$$T_n = \int_0^{\delta} e^{-\lambda y} \cdot y^n \cdot F(-y/E)\,dy = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\lambda y} \cdot y^n \int_{-\infty}^{-y/E} e^{-\frac{t^2}{2}}\,dt\,dy$$

$$T_n = -e^{-\lambda\delta}(\delta)^n F(-\delta/E) + \frac{n}{\lambda} T_{n-1} - \frac{e^{-\lambda^2 E^2/2}}{\lambda} t_n$$

$$\text{and } t_n = \frac{1}{\sqrt{2\pi}} \int_0^{\delta} y^n e^{-\frac{(y+\lambda E^2)^2}{2E^2}}\,dy = \frac{-E}{\sqrt{2\pi}} (\delta)^{n-1} \cdot e^{-\frac{(\delta+\lambda E)^2}{2E^2}} + (n-1)E^2 t_{n-2} - \lambda E^2 t_{n-1}$$

III.2. Double Exponential Distribution

In the previous sections we discussed the gamma distribution. In the following sections double exponential distribution is considered. The variance of all u's is the same and assumed known, u's, the means of the u's may be equal or different. Double exponential is characterized by the density function

$$f(x) = \tfrac{1}{2} \frac{1}{\sqrt{2}\cdot\sigma} e^{-\frac{|x-u|}{\sqrt{2}\cdot\sigma}} \qquad -\infty < x < \infty$$

By virtue of the symmetry round the mean it may be called symmetrized exponential as well. It has mean u and variance σ^2. For algebraic purposes the density function

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

may better be written as

$$f(x) = \tfrac{1}{2}\lambda\, e^{-\lambda|x-u|} \qquad -\infty < x < \infty$$

In this form the mean is u and variance $2/\lambda^2$.

Either of these forms can be obtained from the other by a simple transformation. The absolute sign in the exponential cannot be handled algebraically so it will be in order to remove the absolute sign by noting the form of the density function in different regions.

$$f(x) = \tfrac{1}{2}\lambda\, e^{\lambda u-\lambda x} \qquad x > u$$

$$\left.\begin{aligned} &= \tfrac{1}{2}\lambda e^{-u\lambda+\lambda x} \\ &= \tfrac{1}{2}\lambda e^{-\lambda u-\lambda|x|} \end{aligned}\right\} \tfrac{\lambda}{2}e^{-\lambda u+\lambda x} \qquad 0 < x < u$$

$$x < 0$$

Here u is assumed to be positive. Similar splitting of the density function can be obtained if u is negative but it does not effect the general procedure.

IV. Probability of Correct Selection

In this section we evaluate the probability that the best variety is selected under the assumption that u's themselves follow double exponential distribution. From (2-1) of chapter 2 section I, we have the probability that for given u_i, V_i is selected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$p_i = F_{k-1,\frac{1}{2}}\left[-(u_j-u_i)/E \;;\; j\neq i\right]$$

$$= F_{k-1,\frac{1}{2}}\left(-y_{j/E} \;;\; j\neq i\right) \qquad (3\text{-}16)$$

So the probability depends upon u_i's which are unknown. However they are supposed to follow independently double exponential distribution with known parameter λ and known means U's. All the u's may have different means but in sequel we assume U's (the means of u's) are equal. That will keep the algebra a bit simpler.

As noted above, p_i, the probability that V_i is selected depends upon y_j's which are unknown. For the moment suppose we know the joint density function of y_j ; say $f(y_j \;;\; j\neq i)$. Then the joint probability that V_i is superior to others and is selected is given by integrating (3-16) with respect to $f(y_j; j\neq i)$ over the negative values of y_j and may be written as

$$P_i = \int_{-\infty}^{0}\int_{-\infty}^{0} f(y_j;\; j\neq i)\; F_{k-1,\frac{1}{2}}(-y_j/E \;;\; j\neq i) \prod_{j\neq i} dy_j \qquad (3\text{-}17)$$

Thus the problem reduces to finding the joint density function of y_j. Summing over all i will give the probability that the best variety has been selected. In the particular

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

case when all U's are equal

$$\Sigma P_i = kP_i$$

IV.1.

To find the joint density function of y's two regions are distinguished

(a) $u_i > U$

(b) $u_i < U$

Taking the second case first we treat it for general k and from that can get the density function for any particular k we need. But the first does not lend itself to such a simplification. In that region we have to work for every k separately.

(b) $u_i < U$

For given u_i density function of $u_j - u_i = y_j$ is

$$f(y_j \mid u_i) = \tfrac{1}{2} e^{-\lambda |y_j + u_i - U|}$$

and for given u_i, y_j are statistically independent but they will be ordered. The ordering will depend upon the region and for all y's negative there will be (k-1)! regions, when all U's are equal the regions will be similar. For given u_i density functions of y_j can be multiplied thus giving the joint density function

$$f(y_j \; ; \; j \neq i \mid u_i) = \frac{1}{2^{k-1}} \lambda^{k-1} e^{-\lambda \sum_{j \neq i}^{k} |y_j + u_i - U|} \qquad (3\text{-}18)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Again two cases may be distinguished

(i) $0 < u_i < U$

(ii) $-\infty < u_i < 0$

For negative y's and $0 < u_i < U$

$$f(y_j; u_i \; j\neq i) = \frac{\lambda^k}{2^k} \; e^{-\lambda \sum_{j\neq i} |y_j+u_i-U|} \; e^{-\lambda|u_i-U|}$$

$$= \frac{1}{2^k} \lambda^k \; e^{-\sum_{j\neq i}^{k} \lambda |y_j| \; - \; \lambda k U + \lambda k u_i}$$

Integrating out u_i over the range U to zero we get

$$f(y_j \; ; \; j\neq i) = \frac{\lambda^{k-1}}{k \; 2^k} \cdot \; e^{-\lambda \sum_{j\neq i}^{k} y_j} \; (1 - e^{-\lambda k U}). \qquad (3\text{-}19)$$

For y's negative and $u_i < 0$, noting the form of (3-19) we get the density function

$$f(y_j, \; u_i; \; j\neq i) = \frac{1}{2^k} \lambda^k \; e^{-\lambda \sum_{j\neq i}^{k} |y_j+u_i-U|} \; e^{-\lambda|u_i-U|}$$

$$= \frac{1}{2^k} \lambda^k \; e^{-\lambda \sum_{j\neq i} |y_j| \; - \; \lambda k U - \lambda k \; u_i}$$

Integrating for u_i over the range zero to minus infinity will give the density function

$$f(y_j; \; j\neq i) = \frac{1}{2^k \; k} \; \lambda^{k-1} \; e^{-\lambda \sum_{j\neq i}^{k} |y_j|} \; e^{-\lambda k U} \qquad (3\text{-}20)$$

Thus the sum of the integrated values of (3-16) with respect

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to (3-19) and (3-20) over the negative values of y's will give the joint probability (in the region $u_i < U$) that u_i is the greater than all other u's and V_i is selected. Second part of (3-19) cancels out with (3-20) and thus we are left with

$$P_i^{(1)} = \frac{(k-1)!}{k.2^k} \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \cdots \int_{y_{k-1}}^{\circ} \lambda^{k-1} e^{-\lambda \sum_{j\neq i}^{k} |y_j|} F_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i)\, dy_k \cdots dy_{i+1}\, dy_{i-1} dy$$

(3-21)

where $P_i^{(1)}$ is the probability that $u_i > u_j$ $(j\neq i)$ and V_i is selected for the region $u_i < U$. We may write $P_i = P_i^{(1)} + P_i^{(2)}$ where $P_i^{(2)}$ is defined analogous to $P_i^{(1)}$ for the region $u_i > U$.

It is not hard to see that we would have gotten the same result as (3-21) if U had been taken equal to zero. In fact it is not U's but their differences that really matter. So when we take all U's equal the result is independent of U.

IV.2

In order to determine $P_i^{(2)}$ we take different k separately. Firstly, for k equal to 2

$$f(u_1 - u_2 = y_1 / u_2) = \tfrac{1}{2} e^{-\lambda |y_1 + u_2 - U|} \qquad (3\text{-}22)$$

We work for the region where y_1 is negative and we have $u_2 > U$. Two configurations of y_1 and u_2 arise

(i) $|y_1| > u_2 - U$

(ii) $|y_1| < u_2 - U$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the first of these configurations

$$f(y_1u_2) = \frac{1}{4}\lambda^2 e^{-\lambda|y_1+u_2-U|} e^{-\lambda|u_2-U|}$$

$$= \frac{1}{4}\lambda^2 e^{-\lambda|y_1|+\lambda(u_2-U)-\lambda(u_2-U)}.$$

Integrating for u_2 over the range $U < u_2 < |y_1| + U$ we obtain

$$f(y_1) = \frac{1}{4}\lambda^2 e^{-\lambda|y_1|}|y_1| \qquad (3\text{-}23)$$

for the second configuration

$$f(y_1u_2) = \frac{1}{4}\lambda^2 e^{-\lambda(u_2-U)+\lambda|y_1|} e^{-\lambda(u_2-U)}.$$

Integrating for u_2 over $u_2 > |y_1| + U$ gives

$$f(y_1) = \frac{1}{8}\lambda e^{-\lambda|y_1|} \qquad (3\text{-}24)$$

Thus the sum of the integrated values of (3-16) - for k equal to 2 - with respect to (3-23) and(3-24) over negative values of y_1 will give the probability that $u_{i=2} > u_1$ and V_2 is selected (for the region $u_i > U$) and that will give

$$P_i^{(2)} = \frac{1}{4}\int_{-\infty}^{\circ}\lambda^2 e^{-\lambda|y_1|}|y_1| F(-y_1/E)dy_1 + \frac{1}{8}\lambda\int_{-\infty}^{\circ} e^{-\lambda|y_1|} F(-y/E)dy_1 \qquad (3\text{-}25)$$

Both (3-21) and (3-25) giving $P_i^{(1)}$ and $P_i^{(2)}$ respectively are independent of U. Hence if we can assume that the means of all u's are equal there will be no loss of generality by taking U equal to zero. This should be expected from the symmetry of the specifications. All u's have equal

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

means that leads to the condition that all y's have mean zero. More explicitly in (3-22) and onwards we could substitute $u_2 - U = t$, t has the same distribution as u_2 but its mean is zero. So for higher k we take U equal to zero.

IV.3.

For k equal to 3

Using the result (3-21) we have

$$P_i^{(1)} = \frac{2}{24} \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \lambda^2 . e^{-\lambda|y_1| - \lambda|y_2|} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)\, dy_2 dy_1 \qquad (3\text{-}26)$$

If u_i is greater than zero then there occur six possible configurations of y's and u_i but in the case when all U's are equal they reduce to three basic ones and it will be sufficient to deal with these three only.

The three configurations are

(i) $|y_2| < u_i < |y_1|$

(ii) $u_i < |y_1| < |y_2|$

(iii) $u_i > |y_1| > |y_2|$

For (i) the density function of y_1 and y_2 for given u_i

$$f(y_1, y_2/u_i) = \frac{1}{4} \lambda^2 . e^{\lambda|y_2| - \lambda|y_1| - \lambda(u_i - u_i)}$$

$$= \frac{1}{4} \lambda^2 . e^{-\lambda|y_1| + \lambda|y_2|} \qquad (3\text{-}27)$$

Similarly for the second and the third configurations the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

conditional density functions are

$$\text{(ii)} \qquad \frac{1}{4}\lambda^2 e^{-\lambda|y_1| - \lambda|y_2| + 2\lambda u_i} \tag{3-27-a}$$

$$\text{(iii)} \qquad \frac{1}{4}\lambda^2 e^{+\lambda|y_1| + \lambda|y_2|} e^{-2\lambda u_i} \tag{3-28}$$

The usual procedure of finding the marginal density function of y_1 and y_2 is not available, because u_i itself comes into limits of integration of y_1 and y_2; in other words the limits of y_1 and y_2 are not independent of u_i. This difficulty may be overcome by finding $(P_i^{(2)} / u_i)$ and then integrating that with respect to u_i over the range zero to infinity will give $P_i^{(2)}$. For k equal to three the sum of the integrated values of (3-16) - k = 3 - with respect to (3-26), (3-27-a) and (3-28) - over proper ranges - will give $(P_i^{(2)}/u_i)$ and that integrated with respect to u_i over the range zero to infinity will give $P_i^{(2)}$ and it may be written as

$$P_i^{(2)} = \int_0^\infty (P_i^{(2)}/u_i) f(u_i) du_i$$

$$= 2\Bigg[\int_0^\infty \frac{\lambda^3 e^{-\lambda u_i}}{8} \int_{-u_i}^{0} \int_{-\infty}^{-u_i} e^{-\lambda|y_1| + \lambda|y_2|} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)\, dy_1 dy_2 du_i$$

$$+ \int_0^\infty \frac{\lambda^3 e^{\lambda u_i}}{8} \int_{-\infty}^{-u_i} \int_{-\infty}^{y_1} e^{-\lambda|y_1| - \lambda|y_2|} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)\, dy_2 dy_1 du_i$$

$$+ \int_0^\infty \frac{\lambda^3 e^{-3\lambda u_i}}{8} \int_{-u_i}^{0} \int_{y_1}^{0} e^{+\lambda|y_1| + \lambda|y_2|} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)\, dy_2 dy_1 du_i \Bigg] \tag{3-29}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Multiplication by 2 is on account of the fact that there are two configurations of each type e.g. $|y_2| < u_1 < |y_1|$ and $|y_1| < u_1 < |y_2|$ etc.

IV.4.

For k equal to 4.

From (3-21) we have

$$P_i^{(1)} = \frac{31}{64} \int_{-\infty}^{\circ} \int_{y_1}^{\circ} \int_{y_2}^{\circ} \lambda^3 \, e^{-\lambda|y_1| - \lambda|y_2| - \lambda|y_3|} F_{3,\frac{1}{2}}(-y_j/E; j=1,2,3)\, dy_3 dy_2 dy_1$$

If $u_i > 0$ there are 24 possible configurations of y_1, y_2, y_3 and u_i. For the particular case when all U's are equal they reduce to four basic ones and these four are

(i) $u_i > |y_1| > |y_2| > |y_3|$ and for negative y's

$$f(y_1,y_2,y_3/u_i) = \frac{\lambda^3}{8} e^{+\lambda|y_1| + \lambda|y_2| + \lambda|y_3| - 3u_i\lambda}$$

(ii) $u_i < |y_1| < |y_2| < |y_3|$ with

$$f(y_1,y_2,y_3/u_i) = \frac{\lambda^3}{8} e^{-\lambda|y_1| - \lambda|y_2| - \lambda|y_3| + 3\lambda u_i}$$

(iii) $|y_1| > u_i > |y_2| > |y_3|$ with

$$f(y_1,y_2,y_3/u_i) = \frac{\lambda^3}{8} e^{\lambda|y_2| - \lambda|y_1| + \lambda|y_3| - \lambda u_i}$$

(iv) $|y_1| < u_i < |y_2| < |y_3|$ with

$$f(y_1,y_2,y_3/u_i) = \frac{1}{8} \lambda^3 e^{\lambda|y_1| - \lambda|y_2| - \lambda|y_3| + \lambda u_i}$$

Following the procedure described for k equal to 3 in section IV.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

four - three.

$$P_i^{(2)} = \int_0^\infty (p_i^2/u_i) f(u_i) du_i$$

$$= \frac{6\lambda^4}{16} \Bigg[\int_0^\infty e^{-4\lambda u_i} \int_{-u_i}^{0} \int_{y_1}^{0} \int_{y_2}^{0} e^{+\lambda|y_1| +\lambda|y_2| +\lambda|y_3|} F_{3,\frac{1}{2}}(..) dy_3 dy_2 dy_1 du_i$$

$$+ \int_0^\infty e^{2\lambda u_i} \int_{-\infty}^{-u_i} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} e^{-\lambda|y_1| -\lambda|y_2| -\lambda|y_3|} F_{3,\frac{1}{2}}(..) dy_3 dy_2 dy_1 du_i$$

$$+ \int_0^\infty e^{-2\lambda u_i} \int_{-\infty}^{-u_i} \int_{-u_i}^{0} \int_{y_2}^{0} e^{-\lambda|y_1| +\lambda|y_2| +\lambda|y_3|} F_{3,\frac{1}{2}}(..) dy_3 dy_2 dy_1 du_i \qquad (3\text{-}30)$$

$$+ \int_0^\infty du_i \int_{-u_i}^{0} \int_{-\infty}^{-u_i} \int_{-\infty}^{y_2} e^{+\lambda|y_1| -\lambda|y_2| -\lambda|y_3|} F_{3,\frac{1}{2}}(..) dy_3 dy_2 dy_1 \Bigg]$$

Multiplication by 6 is by virtue of the fact that there are six configurations of each type.

IV.5.

For the present numerical results can be given for k equal to 2 only. For k = 2 and $\lambda_1=\lambda_2=\lambda$ we have

$P = 2P_i$ and P_i is given by the sum of (3-21) for k = 2 and (3-25) and those on simplification give

$$P=2\left| \frac{1}{4} + e^{\frac{\lambda^2E^2}{2}} F(-\lambda E) + \frac{1}{4} \left(\lambda E/\sqrt{2\pi} - e^{\lambda^2E^2/2} \lambda^2E^2 I(\lambda E)\right) \right| \qquad (3\text{-}31)$$

Table 6 gives the probability of correct selection for k equal to 2 for the double exponential a priori distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 6

T	Normal	Double exponential	δ
0	.50	.50	.50
.158	.55	.51-52'	.5422
.3249	.60	.5786	.57271
.5095	.65	.6384	.6227
1.0	.75	.7307	.7727
3.078	.90	.8809	.9903
6.314	.95	.9395	more than .999
31.82	.99	.987	more than .99999

The probabilities in the third column have been calculated on the assumption that the difference between the best variety and the other one is equal to the expected difference between two variates following independently double exponential distribution. And the expected difference (range) for the double exponential variables is $3/2\lambda$. Means of the variables are assumed to be the same.

The values for the double exponential are less than the normal. Comparing with table 1 of chapter 2 it seems that the values for the double exponential are quite similar to the exponential. As long as T is less than one the difference between the third column and the others is not very large but beyond that the values in third column increase over others very rapidly. Similar behaviour was observed in table 1 for the exponential distribution. As in section I.1 of chapter 2,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

T is defined as $T^2 = \frac{n\sigma_o^2}{\sigma^2}$. For the double exponential σ_o^2 is $2/\lambda^2$ and that gives $T = 2/\lambda E$

V. Conditional probability of correct selection

In this section we evaluate the probability of correct selection when the best variety exceeds all others by a specified amount δ and the u's follow independently double exponential distribution. As before probability that V_i is selected

$$= F_{k-1,\frac{1}{2}}(-y_j/E \; ; \; j\neq i) \qquad y_j = u_j - u_i \tag{3-32}$$

Let $f(y_j \; ; \; j\neq i)$ be the joint density function of the y's when all the y's are negative. The conditional probability that V_i is better than others by δ and is selected is given by

$$P_i = \frac{\int_{-\infty}^{-\delta}\cdots\int_{-\infty}^{-\delta} f(y_j;\; j\neq i)F_{k-1,\frac{1}{2}}(-y_j/E \; ; \; j\neq i)\prod_{j\neq i} dy_j}{\int_{-\infty}^{-\delta}\cdots\int_{-\infty}^{-\delta} f(y_j \; ; \; j\neq i)\prod_{j\neq i} dy_j} \tag{3-33}$$

Summing over all i will give the conditional probability of correct selection

$$P = \frac{\sum_{i=1}^{k}\int_{-\infty}^{-\delta}\cdots\int_{-\infty}^{-\delta} f(y_j \; ; \; j\neq i)F_{k-1,\frac{1}{2}}(-y_j/E;\; j\neq i)\prod_{j\neq i} dy_j}{\sum_{i=1}^{k}\int_{-\infty}^{-\delta}\cdots\int_{-\infty}^{-\delta} f(y_j \; ; \; j\neq i)\prod_{j\neq i} dy_j} \tag{3-34}$$

In case the U's (the means of u's) are equal (3-34) will

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

be equivalent to (3-33); its numerator is the joint probability that u_i is greater than all other u's by at least δ and V_i is selected. The denominator is the probability that u_i is greater than all other u's by δ. This simple formulation of the conditional probability when the u's follow double exponential distribution is not available, except for k equal to 2.

V.1.

First of all let us introduce the notation to be used in this section

$$P_i^{\delta} = P_i^{1\delta} + P_i^{2\delta}$$

P_i^{δ} = joint probability that u_i is greater than all u's by δ and V_i is selected.

$P_i^{1\delta}$ = joint probability that u_i is greater than all u's by δ and V_i is selected for the region u_i is less than zero.

$P_i^{2\delta}$ = joint probability that u_i is greater than all other u's by δ when u_i is greater than zero and V_i is selected.

Similarly we define $Q_i^{\delta} = Q_i^{1\delta} + Q_i^{2\delta}$

Q_i^{δ} = the probability that u_i is greater than all other u's by δ.

$Q_i^{1\delta}$ = the probability that u_i is greater than all u's by δ when $u_i < 0$.

$Q_i^{2\delta}$ = the probability that u_i is greater than all other u's by δ when $u_i > 0$.

With this notation the conditional probability will be given by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$P_i = \frac{P_i^{\delta}}{Q_i^{\delta}} = \frac{P_i^{1\delta} + P_i^{2\delta}}{Q_i^{1\delta} + Q_i^{2\delta}}$$

Using the results of (3-21) - for k equal to 2 -, (3-23) and (3-24) we can write the conditional probability of correct selection for k equal to 2 as

$$P=P_i = \frac{\frac{1}{8}\int_{-\infty}^{-\delta} \lambda e^{-\lambda|y|} F(-y/E)dy + \frac{1}{4}\int_{-\infty}^{-\delta} \lambda^2 e^{-\lambda|y|}|y| F(-y/E)dy + \frac{1}{8}\int_{-\infty}^{-\delta} \lambda e^{-\lambda|y|} F(-y/E)d}{\frac{1}{8}\int_{-\infty}^{-\delta} \lambda e^{-\lambda|y|}dy + \frac{1}{4}\int_{-\infty}^{-\delta} \lambda^2 e^{-\lambda|y|}|y|dy + \frac{1}{8}\int_{-\infty}^{-\delta} \lambda e^{-\lambda|y|}dy} \qquad (3\text{-}35)$$

V.2.

For any k, $P_i^{1\delta}$ and $Q_i^{1\delta}$ can be easily written down from (3-21) of section IV.1.

$$P_i^{1\delta} = \frac{(k-1)!}{k\cdot 2^k}\int_{-\infty}^{-\delta}\int_{y_1}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \lambda^{k-1} e^{-\lambda\sum_{j\neq i}^{k}|y_j|} F_{k-1,\frac{1}{2}}(-y_j/E; j\neq i)dy_k..dy_{i+1}dy_{i-1}..dy_1 \qquad (3\text{-}35\text{-a})$$

and

$$Q_i^{1\delta} = \frac{(k-1)!}{k\cdot 2^k}\int_{-\infty}^{-\delta}\int_{y_1}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \lambda^{k-1} e^{-\lambda\sum_{j\neq i}^{k}|y_j|} dy_k..dy_{i+1}dy_{i-1}..dy_1 \qquad (3\text{-}36)$$

In order to find $P_i^{2\delta}$ and $Q_i^{2\delta}$ we have to deal for each k separately. Procedure is indicated for k equal to three.

$$u_i > u_j + \delta \implies u_j - u_i = y_j < -\delta$$

$$u_j - u_i + \delta = y_j + \delta = T_j < 0 \qquad (3\text{-}37)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hence $P_i^{2\delta}$ is the joint probability that T_j's ($j \neq i$) are less than zero and V_i is selected and $Q_i^{2\delta}$ is the probability that T_j are less than zero for the region $u_i > 0$.

We have the density function

$$f(u_j - u_i + \delta = y_j + \delta = T_j / u_i) = \tfrac{1}{2}\, e^{-\lambda |T_j - \delta + u_i|} \qquad T_j < 0$$

For given u_i, T_j are statistically independent so their density functions can be multiplied to get the joint density function. Again for u_i greater than zero six configurations of u_i and $T_j - \delta$ are possible but they reduce to three by symmetry.

(i) $|T_2 - \delta| < u_i < |T_1 - \delta|$ with density function

$$f(T_1 T_2 | u_i) = \tfrac{1}{4}\lambda^2\, e^{-\lambda u_i + \lambda u_i - \lambda |T_1 - \delta| + \lambda |T_2 - \delta|}$$
$$= \tfrac{1}{4}\lambda^2\, e^{\lambda |T_2| - \lambda |T_1|} \qquad (3\text{-}38)$$

In this case u_i is greater than δ because T_1 and T_2 are negative

(ii) $u_i < |T_1 - \delta| < |T_2 - \delta|$ with density function

$$f(T_1 T_2 | u_i) = \tfrac{1}{4}\lambda^2\, e^{-\lambda |T_1| - \lambda |T_2| - 2\lambda\delta + 2\lambda u_i} \qquad (3\text{-}39)$$

(iii) $u_i > |T_1 - \delta| > |T_2 - \delta|$ with the density function

$$f(T_1 T_2 | u_i) = \tfrac{1}{4}\lambda^2\, e^{-2\lambda u_i}\, e^{+\lambda |T_1| + \lambda |T_2| + 2\lambda\delta} \qquad (3\text{-}40)$$

In this case too u_i is greater than δ.

Integrating (3-38), (3-39) and (3-40) - over proper ranges - and summing will give $(Q_i^{2\delta} | u_i)$ and that integrated over relevant range of u_i will give $Q_i^{2\delta}$, thus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$Q_i^{2\delta} = \int (Q_i^{2\delta}/u_i) f(u_i)\, du_i$$

$$= \frac{2\lambda^3}{8} \Bigg[\int_{\delta}^{\infty} e^{-\lambda u_i} \int_{-(u_i-\delta)}^{0} \int_{-\infty}^{-(u_i-\delta)} e^{\lambda|T_2| - \lambda T_1|}\, dT_1 . dT_2\, du_i$$

$$+ \int_{0}^{\infty} e^{\lambda u_i} \int_{-\infty}^{-(u_i-\delta)} \int_{-\infty}^{T_1} e^{-\lambda|T_1| - \lambda|T_2| - 2\lambda\delta}\, dT_2 dT_1 du_i \qquad (3\text{-}41)$$

$$+ \int_{\delta}^{\infty} e^{-3\lambda u_i} \int_{-(u_i-\delta)}^{0} \int_{T_1}^{0} e^{\lambda|T_1| + \lambda|T_2| + 2\lambda\delta}\, dT_2 dT_1 du_i \Bigg]$$

Recalling back the relation (3-37) we have

$$y_j + \delta = T_j$$

or $-|y_j| + \delta = -|T_j|$ defined for $|y_j| \geq \delta$

With this change of variables (3-41) may be written as

$$Q_i^{2\delta} = \frac{2\lambda^3}{8} \Bigg[\int_{\delta}^{\infty} e^{-\lambda u_i} \int_{u_i}^{-\delta} \int_{-\delta}^{-u_i} e^{-\lambda|y_1| + \lambda|y_2|} dy_1 dy_2 du_i$$

$$+ \int_{0}^{\infty} e^{\lambda u_i} \int_{-\infty}^{-u_i} \int_{-\infty}^{y_1} e^{-\lambda|y_1| - \lambda|y_2|}\, dy_2 dy_1 du_i$$

$$+ \int_{\delta}^{\infty} e^{-3\lambda u_i} \int_{-u_i}^{-\delta} \int_{y_1}^{-\delta} e^{\lambda|y_1| + \lambda|y_2|}\, dy_2 dy_1 du_i \Bigg] \qquad (3\text{-}42)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(3-32) for k equal to three integrated with respect to (3-42) will give $P_i^{2\delta}$; - the joint probability that u_i is greater than all u's by δ and V_i is selected and may be written as

$$P_i^{2\delta} = \int Q_i^{2\delta} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)dy_1dy_2 \qquad (3\text{-}43)$$

and we have the probability of correct selection

$$P = P_i^{\delta} = \frac{P_i^{1\delta} + P_i^{2\delta}}{Q_i^{1\delta} + Q_i^{2\delta}}$$

Numerator is given by the sum of (3-35-a) for k equal to 3 and (3-43) and the denominator is given by the sum of (3-36) and (3-42). The procedure can be extended to higher values of k in a straightforward manner but the numerical calculation will be extremely unwieldy.

In chapter 2 as well as in the present chapter the probability values for different a priori distributions have been compared with what has come to be known as "least favourable configuration". And seemingly paradoxically, least favourable configuration always gives higher values for the probability of correct selection. This apparent paradox is due to the words used rather than any real inconsistency. When we talk of least favourable configuration it is implicitly understood that least favourable is with respect to a specified δ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the a priori distribution assumption we allow any variety to be arbitrarily close to the best one but, that is not possible in what is termed as least favourable configuration.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4

NUMERICAL EVALUATIONS

The mathematical formulae have their own intrinsic value but, to be of real use it should be possible to have their numerical values. Dunnett (19) has given numerical results for k equal to 2 for the normal case. We start with the exponential. In chapter 2 two alternative procedures were given for the exponential a priori distribution. If λE $(\frac{E}{\sigma_o})$ is such that the probability of correct selection is fairly high, say 90% and that should cover most of the practical situations, by combining the two procedures fairly good approximation can be obtained for all k. The approximate method has been adopted for double exponential and results can be obtained for $k \leq 4$, beyond that the accuracy is very poor. Then the method is adopted for Dunnett's (19) formulae and by that it has been possible to get numerical results for k equal to 3 for the normal. For the gamma for the time being it has not been possible to get numerical results for k greater than 2.

I. Exponential

Reproducing the formulae from section I of chapter 2 we have

$$\text{Prob (for given } u_i, V_i \text{ is selected)} = F_{k-1,\frac{1}{2}}(-y_j/E;\ j \neq i) \qquad (4\text{-}1)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Region R_1 has been defined there as

$$|y_1| > |y_2| > \cdots |y_{i-1}| > |y_{i+1}| > \cdots > |y_k|$$

and for this region probability that u_i is greater than all other u's is given as

$$\int_{-\infty}^{0}\int_{-\infty}^{y_{k-1}}\cdots\int_{-\infty}^{y_2} \frac{\prod_{j=1}^{k}\lambda_j}{\sum_{i=1}^{k}\lambda_i} e^{-\sum_{j\neq i}\lambda_j y_j + y_i\sum_{j=1}^{k}\lambda_j} dy_1 \cdots dy_{i-1}\, dy_{i+1}\cdots dy_k \tag{4-2}$$

These two formulae give the probability that u_i is the largest and V_i is selected for the region R_1

$$R_{1i} = \int_{-\infty}^{0}\int_{-\infty}^{y_{k-1}}\cdots\int_{-\infty}^{y_2} \frac{\prod_{j=1}^{k}\lambda_j}{\sum_{i=1}^{k}\lambda_i} e^{y_i\sum_{j=1}^{k}\lambda_j - \sum_{j\neq i}\lambda_j y_j} F_{k-1,\frac{1}{2}}(-y_j/E;\, j\neq i)\, dy_1 \cdots dy_{i-1}\, dy_{i+1}\cdots dy_k \tag{4-3}$$

So our task is reduced to find a solution for (4-3) for general k; (4-3) gives the joint probability that $u_i > u_j$ and $\bar{X}_i > \bar{X}_j$ for all $j\neq i$ for the region R_1 and is designated by R_{1i}. Going through similar steps we can get R'_{1i} = Prob ($u_i > u_j$ and $\bar{X}_i < \bar{X}_j$ for all $j\neq i$). In other words R'_{1i} is the probability that u_i is greater than all other u's but V_i happens to be the lowest in sample means.

$$R'_{1i} = \int_{-\infty}^{0}\int_{-\infty}^{y_{k-1}}\cdots\int_{-\infty}^{y_2} \frac{\prod_{j=1}^{k}\lambda_j}{\sum_{i=1}^{k}\lambda_i} e^{-\sum_{j\neq i}\lambda_j y_j + y_i\sum_{j=1}^{k}\lambda_j} F_{k-1,\frac{1}{2}}(y_j/E;\, j\neq i)\, dy_1 \cdots dy_{i-1}\cdots dy_{i+1}\cdots dy_k \tag{4-4}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Definition: A_j is the event that $u_i > u_j$ for all $j \neq i$ and $\bar{X}_i < \bar{X}_j$, $(j \neq i)$

Thus defined A_1 is the event that u_i is greater than all other u's and sample mean of V_i is less than the sample mean of V_1. Similarly are defined the events A_2, A_3..etc. and there will be k-1 of them.

$$\Pr(A_1+A_2+..+A_{i-1}+A_{i+1}+..+A_k) = \text{Prob(that at least one of the A's occur)}$$

$$= P(u_i > u_j \text{ for all } j \neq i \text{ and } \bar{X}_i < \bar{X}_j \text{ for at least one } j)$$

From (4-2) we have prob $(u_i > u_j$ for all $j \neq i) = \dfrac{\lambda_i}{(k-1)! \sum_{j=1}^{k} \lambda_j}$

$$= \frac{1}{k(k-1)!} \text{ for all } \lambda\text{'s equal.}$$

From now on we take all λ's equal. If they are different the procedure will still work though the algebraic work will increase k times. From this we have

$$P(A_1+A_2+..+A_{i-1}+A_{i+1}+..+A_k)$$

$$= \frac{1}{k!} - \text{Prob } (u_i > u_j \text{ for all } j \neq i \text{ and } \bar{X}_i < \bar{X}_j \text{ for no } j: j \neq i)$$

$$= \frac{1}{k!} - \text{Prob } (u_i > u_j \text{ and } \bar{X}_i > \bar{X}_j \text{ for all } j \neq i) \qquad (4\text{-}5)$$

Second term on the right hand side of (4-5) is the same as R_{1i} in (4-3). Transposing the terms we get

$$R_{1i} = \frac{1}{k!} - P\ (A_1+A_2+\ldots\ A_{i-1}+A_{i+1}+..A_k)$$

$$= \frac{1}{k!} - (\Sigma\ A_j - \Sigma\ A_jA_k+\ldots.\ )$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Denoting the event and the probability of the event by the same letter)

Thus the problem reduces to find the A_j etc.

As λ's are equal so the numbering of A's can be arbitrary. For k equal to three there will be two A's say A_1 and A_2

$$A_2 = \int_{-\infty}^{\circ}\int_{-\infty}^{y_2} \frac{\lambda^2}{3} e^{\lambda y_1 + \lambda y_2} F\left(\frac{y_2}{E}\right) dy_2 dy_2 = \frac{1}{6}\left[\frac{1}{2} - e^{\frac{\lambda^2 E^2}{2}} I(\lambda E)\right]$$

$$A_2 = \int_{-\infty}^{\circ}\int_{-\infty}^{y_2} \frac{\lambda^2}{3} e^{\lambda y_1 + \lambda y_2} F\left(\frac{y_2}{E}\right) dy_2 dy_2 = \frac{1}{6}\left[\left(\frac{1}{2} - e^{\frac{\lambda^2 E^2}{2}} I(\lambda E)\right.\right.$$

$$\left. + e^{2\lambda^2 E^2} I(2\lambda E) - e^{\frac{\lambda^2 E^2}{2}} I(\lambda E)\right]$$

$$A_1 A_2 = \frac{1}{6}\left[F_{2,\frac{1}{2}}(o,o) + e^{2\lambda^2 E^2}\int_{-\infty}^{-2\lambda E}\int_{-\infty}^{-\lambda E} f_{2,\frac{1}{2}}\, dx\, dy - 3e^{\frac{\lambda^2 E^2}{2}}\int_{-\infty}^{-\lambda E}\int_{-\infty}^{y_2+\frac{\lambda E}{2}} f_{2,\frac{1}{2}}\, dx\, dy_2\right.$$

$$\left. - e^{\lambda^2 E^2}\int_{-\infty}^{\frac{3}{2}}\int_{-\infty}^{\circ} f_{2,\frac{1}{2}}\, dx\, dy_2\right] \qquad (4\text{-}5)$$

A_1 and A_2 involve univariate integrals only but A_1A_2 involves bivariate integrals. The first two terms of A_1A_2 are very easy to evaluate, tables are available for that. Cadwell (20) has given a procedure to evaluate the integrals of type of the last two terms of A_1A_2. For k equal to four, there will be three A's, say, A_1, A_2 and A_3. These can be found in a straightforward manner and they involve univariate integrals only. The terms like A_1A_2 etc. will include terms like those in (4-5) but there will be about 7 of them. Thus their

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluation will not be very easy. At this level the procedure described in section IV of chapter 2 may be made use of. This point may be illustrated with the help of the following two tables. Table 7 gives the probability of correct selection for k equal to 2 and 3.

Table 7

k \ λE	.05	.1	.2	.3	.4	.5	.75	1.0	2.0	2.6
2	.9807	.9624	.9296	.8996	.8733	.8497	.8002	.7616	.6683	.6372
3	.9796	.9597	.9151	.874	.831	.796	.715	.605		

In section IV.1. of chapter 2 we derived the expression which gives the probability that rth best variety exceeds the best one and it is

$$(r-1)\int_0^\infty (1-e^{-\lambda y})^{r-2}\, e^{-\lambda y}\, F(-\tfrac{y}{E})\,dy \qquad (4\text{-}6)$$

From this for any 'r' we can evaluate the probability and table 8 gives some of the values.

Table 8

Probability that the rth variety exceeds the first

r \ λE	.05	.1	.2	.3	.4	.5	.75	1.0	2.0	2.6
2	.0193	.03751	.07074	.10036	.12669	.15035	.19976	.23836	.33177	.03628
3	.00108	.00408	.0188	.0290	.0466	.0623	.1056	.20494	.2581	
4	.000002	.0006	.0036	.0102	.0234	.0327	.0639			
5				.0042	.0092	.0148				

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From table 8 it is clear that for that λE for which the probability of correct selection will be 90% or over the probability of rth ($r \geq 4$) variety exceeding the 1st is fairly small. Table 7 combined with table 8 will give for such λE values correct up to 2 decimal places for all k. The way to combine the two tables is explained below.

Example: Find the probability of correct selection for $k=k_1$ for $\lambda E=0.1$. From chapter 2 we have

$$\text{Prob(correct selection)} = 1 - \text{Prob}(\bar{X}_2 > \bar{X}_1) - \text{Prob}(\bar{X}_3 > \bar{X}_1 > \bar{X}_2) - \cdots$$

$$- \text{Prob}\left[(\bar{X}_{k_1} > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{k_1-1}) \right]$$

We also know that for the exponential all the terms are independent of k. Thus we can write the probability of correct selection step by step.

For $k = k_1 + 1$

$$\text{Prob(correct selection} = \text{Prob(correct selection for } k=k_1)$$

$$- \text{Prob}\left[(\bar{X}_{k_1+1} > \bar{X}_1 > (\bar{X}_2 \ldots \bar{X}_{k_1}) \right]$$

From table 7 we have for $\lambda E = 0.1$

Prob (correct selection for $k = 3$) = .9597

From table 8 we have the upper limit of $P\left[\bar{X}_4 > \bar{X}_1 > (\bar{X}_2, \bar{X}_3)\right] = .0036$

Therefore Prob(correct selection for $k=4$) $> .9597 - .0036$

The process can be continued for higher k but it is evident from table 8 that further terms will be so small that they can be ignored.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

II. Double Exponential

For the double exponential exact solution can be obtained for k equal to three; even that involves the evaluation of about twenty double integrals. For the integrals the regions over which they are to be evaluated are not rectangles (for which there is no problem, tables are available) but polygons; that makes the exact solution very near impossible. The approximate method described above can be used to get some approximation and it will give the lower limit of the probability value.

Probability P_i that u_i is greater than all u's and V_i is selected for the double exponential is defined in section IV.1 chapter 3 following equation (4-21) as $P_i = P_i^{(1)} + P_i^{(2)}$; $P_i^{(1)}$ and $P_i^{(2)}$ are also defined there. For k equal to three $P_i^{(1)}$ is given by (3-26) of section IV.3 chapter 3. Apart from constants its evaluation is exactly similar to exponential distribution. $P_i^{(2)}$ is given in (3-29) of section IV.3 chapter 3 and it has three constituents. We give the solution for the first of these three; for the other two it will be a straightforward application. First part of $P_i^{(2)}$, let us call it $P_{i_1}^{(2)}$;

$$P_{i_1}^{(2)} = \int_0^\infty \frac{\lambda^3 e^{-\lambda u_i}}{8} \int_{-u_i}^{0} \int_{-\infty}^{-u_i} e^{-\lambda|y_1| + \lambda|y_2|} F_{2,\frac{1}{2}}(-y_1/E, -y_2/E)\, dy_1 dy_2 du_i \qquad (4\text{-}7)$$

$P_{i_1}^{(2)}$ is the joint probability that u_i is greater than u_1 and u_2 (for the region $u_i > 0$ and the configuration $|y_2| < u_1 < |y_1|$) and V_i is selected. For this configuration

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the probability that u_i is greater than u_1 and u_2 is given by multiplying the conditional density function of y_1 and y_2 given in (3-27) by the density function of u_i and integrating it over proper ranges and it is

$$\int_0^\infty \frac{\lambda^3 e^{-\lambda u_i}}{8} \int_{-u_i}^{0} \int_{-\infty}^{-u_i} e^{-\lambda|y_1|+\lambda|y_2|}\, dy_1 dy_2 = \frac{1}{16} \tag{4-8}$$

Define A_j the event that $u_i > u_j$ and $\bar{X}_i < \bar{X}_j$ $j=1$ and 2

That will give $P(A_1+A_2)$

$$= \text{Prob}(u_i > u_j \text{ and } \bar{X}_i < \text{at least one of } \bar{X}_j,\ j=1 \text{ and } 2)$$

$$= \frac{1}{16} - P(u_i > u_j \text{ and } \bar{X}_i > \bar{X}_j \quad j=1 \text{ and } 2) \tag{4-9}$$

In (4-9) second part of right hand side is equal to $P_{i_1}^{(2)}$; thus

$P_{i_1}^{(2)} = \frac{1}{16} - P(A_1+A_2) = \frac{1}{16} - P(A_1) - P(A_2) + P(A_1A_2)$.

A_1 and A_2 can be evaluated easily and $\left[\frac{1}{16} - P(A_1) - P(A_2)\right]$ will give the lower limit of $P_{i_1}^{(2)}$

$$A_1 = \frac{1}{8}\lambda^3 \int_0^\infty e^{-\lambda u} \int_{-u}^{0} \int_{-\infty}^{-u} e^{-\lambda|y_1|+\lambda|y_2|} F(y_1/E)\, dy_1 dy_2 du \tag{4-10}$$

$$= \frac{1}{8}\left[\frac{1}{4} + \lambda^2E^2 e^{\frac{\lambda^2E^2}{2}} I(\lambda E) - \frac{\lambda E}{\sqrt{2\pi}} - \tfrac{1}{2}e^{\lambda^2E^2\cdot 2} I(2\lambda E)\right]$$

$$A_2 = \frac{1}{8}\left[\frac{1}{4} - \frac{e^{\lambda^2E^2\ 2}}{2} I(\lambda E)\right]$$

The same procedure can be extended to k equal to 4.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For k equal to 4, $P_i^{(1)}$ is given in section IV.3 of chapter 3. Its evaluation is just like the exponential. Here $P_i^{(2)}$ consists of four parts, the first part - denote it by $P_{i_1}^{(2)}$;

$$P_{i1}^{(2)} = \frac{\lambda 4}{16} \int_0^\infty e^{-4\lambda u} \int_{-u}^{\circ} \int_{y_3}^{\circ} \int_{y_2}^{\circ} e^{\lambda(|y_1|+|y_2|+|y_3|)} F_{3,\frac{1}{2}}(-y_j/E; j=1,2,3)\, dy_1 dy_2 dy_3 du$$

(4-11)

Defining A's like those for k equal to 3 we will get $P_{i1}^{(2)} = \frac{1}{96} - P(A_1+A_2+A_3)$; thus $\frac{1}{96} - P(A_1)-P(A_2)-P(A_3)$ will give the lower limit of $P_{i1}^{(2)}$

$$A_1 = \frac{\lambda^4}{16} \int_0^\infty e^{-4\lambda u} \int_{-u}^{\circ} \int_{y_3}^{\circ} \int_{y_2}^{\circ} f(y_1y_2y_3) F(y_1/E)\, dy_1 dy_2 dy_3 du$$

$$= \frac{1}{16} \left[\frac{1}{48} - \frac{1}{8} e^{\frac{\lambda^2E^2}{2}} I(\lambda E) + \frac{1}{8} e^{2\lambda^2E^2} I(2\lambda E) - \frac{e^{\frac{9}{2}\lambda^2E^2}}{24} I(3\lambda E) \right]$$

Likewise

$$A_2 = \frac{1}{16} \left[\frac{1}{48} - \frac{e^{\lambda^2E^2/2}}{12} I(\lambda E) + \frac{e^{2\lambda^2E^2}}{24} I(2\lambda E) \right]$$

$$A_3 = \frac{1}{16} \left[\frac{1}{48} - \frac{e^{\lambda^2E^2/2}}{24} I(\lambda E) \right] \qquad (4\text{-}12)$$

III. Normal:

Dunnett (19) has worked with normal distribution but he has been able to give numerical results for k equal to 2 only. The above procedure can be adopted for this case also to extend the numerical results for k equal to 3.

For k=3 Dunnett's (19) derivation will be somewhat as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

follows. Probability of any variety, say V_3 to be selected is

$$= F_{2,\frac{1}{2}} \left[-(u_2-u_3)/E, \; -(u_1-u_3)/E \right]$$

$$= F_{2,\frac{1}{2}} (-y_1/E, -y_2/E) \qquad (4\text{-}13)$$

But u's follow independently normal distribution with common variance σ_o^2 and known means U_1, U_2 and U_3. From this stipulation we get the probability that u_3 is greater than u_1 and u_2

$$p_3 = \int_{-\infty}^{-(U_2-U_3)/\sqrt{2}\sigma_o} \int_{-\infty}^{-(U_1-U_3)/\sqrt{2}\sigma_o} f_{2,\frac{1}{2}} dy_1 dy_2 \qquad (4\text{-}14)$$

where $f_{2,\frac{1}{2}}$ is a bivariate density function with correlation $\frac{1}{2}$. The joint probability that u_3 is greater than u_1 and u_3 and V_3 will be selected will be given by integrating (4-13) with respect to (4-14). That will be a four variate integral.

Similar to section II define A_1 and A_2.

Definition: A_1 is the event that u_3 is greater than u_1 and u_2 and $\bar{X}_3 < \bar{X}_1$

A_2 is the event that u_3 is greater than u_1 and u_2 and $\bar{X}_3 < \bar{X}_2$

Going through similar arguments as in section I we get $P_{i=3} = p_3 - P(A_1+A_2)$ and $p_3 - P(A_1) - P(A_2)$ will give the lower limit of $P_{i=3}$. So the problem is reduced to find $P(A_1)$ and $P(A_2)$.

Probability that $\bar{X}_3 < \bar{X}_1 = F^{\frac{(u_1-u_3)}{E}} = F(y_1/E)$ (4-15)

$P(A_1)$ will be given by integrating (4-15) with respect to (4-14).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$P(A_1) = \int_{-\infty}^{-(U_2-U_3)/\sqrt{2}\sigma_o} \int_{-\infty}^{-(U_1-U_3)/\sqrt{2}\sigma_o} f_{2,\frac{1}{2}}(y_1,y_2)F(y_1/E)dy_1dy_2 \qquad (4\text{-}16)$$

This is a trivariate normal integral with correlation structure

$$\rho(y_1y_2) = \tfrac{1}{2}$$
$$\rho(y_2y_3) = -\tfrac{1}{2}\rho \qquad \rho = \frac{\sqrt{2}\sigma_o}{\sqrt{2\sigma_o^2+E^2}}$$
$$\rho(y_1y_3) = -\rho$$

The limits of integration for two variates are given and for the third it will be zero to minus infinity. If all U's are equal the range of integration will be zero to minus infinity for all the three variables and $P(A_1)$ will reduce to

$$\int_{-\infty}^{0}\int_{-\infty}^{0}\int_{-\infty}^{0} f(y_1y_2y_3)dy_1dy_2dy_3 \qquad (4\text{-}17)$$

and by symmetry $P(A_2)$ will have the same value. And with all U's equal p_3 of (4-14) will have the value of .3333. Solution of (4-17) is well known. It is a generalization of Sheppard's (31) formula for the bivariate normal integral

$$(5\text{-}17) = \Phi(o,o,o;\rho_{12},\rho_{13},\rho_{23}) = \frac{1}{4\pi}(2\pi - \arccos\rho_{12} - \arccos\rho_{23} - \arccos\rho_{13})$$

Table 9 gives the probabilities of correct selection for k equal to 3 for exponential, double exponential and normal distributions. Only such values of λE $(\frac{E}{\sigma_o})$ have been taken which give pretty high values for the probability.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the double exponential and the normal only lower limits are given. For k equal to three exponential gives the least values and the double exponential and the normal give very similar values.

Table 9

Probability of correct selection k=3

$\lambda E = \frac{\lambda}{\sigma_o}$	Exponential	Double Exponential L.L.	Normal L.L.
.05	.9796	.9849	.9928
.1	.9591	.9695	.9707
.2	.9151	.93822	.9302
.3	.874	.90710	.8943

Table 9.a

Probability of correct selection k=4

$\lambda E = \frac{E}{\sigma_o}$	Exponential L.L.	Double Exponential L.L.
.05	0.9796	.9681
.1	0.9585	.9444
.2	0.9115	.89178

From the above tables it is clear that for k up to three double exponential gives higher values than the exponential but for k equal to four the position is reversed. This may be explained as follows:

The expected range of two variates following exponential distribution is $1/\lambda$, which is equal to the standard deviation; in the double exponential case the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

expected range is $\frac{3}{2\lambda}$, which is greater than the standard deviation. As k is increased the expected difference of the best and the 2nd best remains the same for the exponential where as in the double exponential it decreases.

IV. Exponential - conditional probability

The procedure outlined earlier can be adopted to get the conditional probability of correct selection also. Conditional probability of correct selection for exponential is given in (2-16) of chapter 2. When all λ's are equal the summation in the numerator as well as in the denominator will be replaced by multiplication and it mutually cancels out. The denominator can be evaluated in a straightforward manner. The numerator is the joint probability that V_i is at least δ superior to all others and it is selected. Going through similar reasoning we will get the probability that V_i is at least δ superior to all others and it appears as inferior to all others as

$$\int_{-\infty}^{-\delta}\int_{y_1}^{-\delta}\cdots\int_{y_{k-1}}^{-\delta} \frac{\prod_{j=1}^{k}\lambda_j\, e^{\sum_{j\neq i}\lambda_j y_j}}{\sum_{j=1}^{k}\lambda_j}\, e^{y_1\sum_{i=1}^{k}\lambda_i} F_{k-1,\frac{1}{2}}(y_j/E;\, j\neq i)\, dy_k \cdots dy_{i+1}\, dy_{i-1}\, dy_1 \qquad (4\text{-}18)$$

Similar to section I. define A_j, as the event that $u_i > u_j$ by δ for all $j\neq i$ and $\bar{X}_i < \bar{X}_j$. Thus defined A_1 is the event that u_i is greater than all other u's by δ and sample mean of V_i is less than the sample mean of V_I. Similarly, are defined the events A_2, A_3 and there will be k-1 of them.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$Pr(A_1+..+A_{i-1}+A_{i+1}..+A_k)$

= Prob(that at least one of the A's occur)

= Prob($u_i > u_j+\delta$ for all $j\neq i$ and $\bar{X}_i < \bar{X}_j$ for at least one j)

The denominator of (4-18) gives the probability that $u_i > u_j + \delta$ for all $j\neq i$. For given δ it can be worked in a straightforward manner, let it be denoted by p. From that we get $P(A_1+A_2+..+A_{k-1})$

$$=p - \text{Prob}(u_i > u_j+\delta \text{ for all } j\neq i \text{ and } \bar{X}_i < \bar{X}_j \text{ for no } j:)$$

$$=p - P(u_i > u_j+\delta \text{ for all } j\neq i \text{ and } \bar{X}_i > \bar{X}_j \text{ for all } j\neq i).$$

Transposing we will get

$$\text{Prob}(u_i > u_j+\delta \text{ and } \bar{X}_i > \bar{X}_j \text{ for all } j\neq i) = p - P(A_1+..+A_{i-1}+A_{i+1}+..+A_k)$$

Thus the problem reduces to find $P(A_j)$ etc. and they can be determined just like those explained in section I.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5

MISCELLANEOUS EXTENSIONS

So far we have considered three problems, (a) probability of correct selection under all circumstances, (b) conditional probability of correct selection and (c) that the selected variety is within δ neighbourhood of the maximum. None of these formulations bring the economic aspect into picture. Perhaps δ (in conditional probability etc.) would be fixed by the experimenter by taking into consideration the economic aspect of the experiment. The probability values (in all three forms) can be arbitrarily increased by increasing the size of the experiment but, that will result in increased costs. So some sort of balance has to be obtained. In these sections we discuss a procedure to arrive at the optimum size of the experiment under the condition that the varietal means follow independently exponential distribution. Every observation on each variety is normally distributed with known variance. The procedure is derived for k equal to 2 only. The optimum size is in terms of λ and δ. There we take the case where u's follow gamma distribution. For gamma r is taken equal to 2 only. Though the method is straightforward for higher r the algebra becomes pretty tedious. For k greater than 2 the method can be extended but that will require the values of multiple integrals and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

algebra becomes extremely unweildy.

I. For the problem of optimum size of the experiment we need certain results and those are derived in this section.

(a) x_1 and x_2 follow independently normal distribution with common known variance and means θ_1 and $\theta_2 \neq \theta_1$ and θ_2 in turn follow exponential distribution with parameters λ_1 and λ_2; determine the expected value of the x_{max}.;

for x_{max} the density function is

$$f(x_{max}) = \frac{1}{\sigma} f\left(\frac{x-\theta_1}{\sigma}\right) F\left(\frac{x-\theta_2}{\sigma}\right) + \frac{1}{\sigma} f\left(\frac{x-\theta_2}{\sigma}\right) F\left(\frac{x-\theta_1}{\sigma}\right) .$$

$$-\infty < x < \infty$$

that will give

$$E(x_{max}) = \frac{1}{\sigma} \int_{-\infty}^{\infty} x f\left(\frac{x-\theta_1}{\sigma}\right) F\left(\frac{x-\theta_2}{\sigma}\right) dx + \int_{-\infty}^{\infty} \frac{x}{\sigma} f\left(\frac{x-\theta_2}{\sigma}\right) F\left(\frac{x-\theta_1}{\sigma}\right) dx \qquad (5\text{-}1)$$

only the first term of (5-1) is evaluated.

$$T = \frac{1}{\sigma} \int_{-\infty}^{\infty} x f\left(\frac{x-\theta_1}{\sigma}\right) F\left(\frac{x-\theta_2}{\sigma}\right) dx$$

$$= \int_{-\infty}^{\infty} \frac{(x-\theta_1)}{\sigma} f\left(\frac{x-\theta_1}{\sigma}\right) F\left(\frac{x-\theta_2}{\sigma}\right) dx + \frac{\theta_1}{\sigma} \int_{-\infty}^{\infty} f\left(\frac{x-\theta_1}{\sigma}\right) F\left(\frac{x-\theta_2}{\sigma}\right) dx$$

. . . . . .

$$= \frac{\sigma}{2\sqrt{\pi}} e^{-\frac{(\theta_1^2+\theta_2^2-2\theta_1\theta_2)}{4\sigma^2}} + \frac{\theta_1}{\sqrt{2\pi}} \int_{-\infty}^{(\theta_1-\theta_2)/\sqrt{2}\cdot\sigma} e^{\frac{-t^2}{2}} dt \qquad (5\text{-}2)$$

But θ_1 and θ_2 are supposed to follow exponential distribution with parameters λ_1 and λ_2. Thus integrating (5-2) with respect to $\lambda_1\lambda_2\, e^{-\lambda_1\theta_1-\lambda_2\theta_2}$ will give the desired expected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

value of max x; straightforward algebra gives the integrated value as

$$\frac{2\lambda_1\lambda_2}{(\lambda_1+\lambda_2)} \cdot \sigma^2 \cdot e^{\lambda_1^2\sigma^2} I(\sqrt{2}\cdot\sigma\lambda_1) + \frac{1}{2\lambda_1}$$

$$+ \frac{1}{\lambda_1} e^{\lambda_1^2\sigma^2} I(\sqrt{2}\,\sigma\lambda_1) + \frac{\sigma}{\sqrt{\pi}} - 2\sigma^2\lambda_1 \cdot e^{\lambda_1^2\sigma^2} I(\sqrt{2}\,\sigma\lambda_1) - \frac{2\lambda_1}{(\lambda_1+\lambda_2)^2} e^{\sigma^2\lambda_2^2} I(\sqrt{2}\sigma\lambda_1)$$

$$- \frac{\sigma}{\sqrt{\pi}} \cdot \frac{\lambda_1}{\lambda_1+\lambda_2} + e^{\sigma^2\lambda_1^2} \frac{2\,\sigma^2\lambda_1^2}{(\lambda_1+\lambda_2)} I(\sqrt{2}\cdot\sigma\lambda_1).$$

If we take $\lambda_1=\lambda_2=\lambda$ the above expression simplifies to

$$\frac{1}{2\lambda} + \frac{\sigma}{2\sqrt{\pi}} + \frac{e^{\lambda^2\sigma^2}}{2\lambda} I(\sqrt{2}\cdot\sigma\lambda) \qquad (5\text{-}3)$$

As σ approaches zero (5-3) approaches $3/4\lambda$ and that is what can be expected. When σ is zero the variables x_1 and x_2 are identified with θ_1 and θ_2 which are supposed to follow exponential distribution with parameter λ and hence the expected maximum is the same as the expected maximum of two variables following exponential distribution and that is (first term only)

$$\lambda^2 \int_0^\infty xe^{-\lambda x} \int_0^x e^{-\lambda t}\, dt\, dx = 3/4\lambda$$

If λ approaches infinity then (5-3) attains the value $\sigma/2\sqrt{\pi}$; that is what can be obtained by substituting zero for θ_1 and θ_2 in (5-2) because when λ approaches infinity mean of x_1 and x_2 will approach zero and expected max. x will be maximum of two normally distributed variables with zero means and common variance σ^2. By symmetry the other term will be of the same value. Thus under the conditions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

we get

$$E(x_{max}) = \frac{1}{\lambda} + \frac{\sigma}{\sqrt{\pi}} + \frac{e^{\sigma^2\lambda^2}}{\lambda} \quad I(\sqrt{2}\sigma\lambda) \qquad (5\text{-}4)$$

Going through the same arguments and similar algebra we can get $E(x_{max})$ for the gamma distribution with r equal to 2 (First term only)

$$\frac{1}{\lambda} + \frac{3\sigma}{4\sqrt{\pi}} + \frac{3}{4}\frac{e^{\lambda^2\sigma^2}}{\lambda} I(\sqrt{2}\sigma\lambda) - \frac{\sigma}{2\sqrt{2}} \sqrt{2}\, e^{\lambda^2\sigma^2} I(\sqrt{2}\sigma\lambda) \qquad (5\text{-}5)$$

As σ approaches zero (5-5) approaches $\frac{1}{\lambda} + \frac{3}{8\lambda} = \frac{11}{8\lambda}$ and that is again by virtue of x_1 and x_2 being identified with θ_1 and θ_2 which are gamma variables. As λ approaches infinity the above expression approaches $\frac{\sigma}{2\sqrt{\pi}}$ i.e. by substituting zero in (5-2). By virtue of symmetry the other term will be of the same value as (5-5), so for gamma with r equal to 2

$$E(x_{max}) = \frac{2}{\lambda} + \frac{3\sigma}{2\sqrt{\pi}} + \frac{3}{2\lambda} e^{\lambda^2\sigma^2} I(\sqrt{2}\sigma\lambda) - \frac{\sigma}{\sqrt{2}} \sqrt{2}\cdot\lambda\sigma\, e^{\lambda^2\sigma^2} I(\sqrt{2}\sigma\lambda) \qquad (5\text{-}6)$$

(b) x is a normal variable with mean θ and variance σ^2 ; θ is supposed to follow exponential distribution with parameter λ ; to find the regression of θ on x.

$$f(x) = \frac{1}{\sqrt{(2\pi)}} e^{-(x-\theta)^2/2\sigma^2}$$

$$f(x\theta) = \frac{1}{\sqrt{(2\pi)}} e^{-(x-\theta)^2/2\sigma^2} \cdot \lambda . e^{-\lambda\theta}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from this we get $E(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x\, e^{-(x-\theta)^2/2\sigma^2} \int_{0}^{\infty} \lambda\, e^{-\lambda\theta} dx\, d\theta = 1/\lambda$

$$E(\theta) = 1/\lambda$$

$$E(x\theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x\theta\, e^{-(x-\theta)^2/2\sigma^2} \int_{0}^{\infty} \lambda\, e^{-\lambda\theta} dx\, d\theta$$

$$= 2/\lambda^2$$

Thus covariance of $(x\theta) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$

$$E(x^2) = \frac{1}{\sqrt{2\pi}.\,\sigma} \int_{-\infty}^{\infty} x^2\, e^{-(x-\theta)^2\, 2\sigma^2} \int_{0}^{\infty} \lambda\, e^{-\lambda\theta} dx\, d\theta = \sigma^2 + 2/\lambda^2$$

That will give $var(x) = \sigma^2 + 1/\lambda^2$ and from that regression of θ on x is obtained

$$= \frac{1/\lambda^2}{\sigma^2 + 1/\lambda^2} = \frac{1}{\sigma^2\lambda^2 + 1} \qquad (5\text{-}7)$$

Going through similar algebra we get the regression of θ on x when θ is gamma as $\dfrac{1}{\frac{\sigma^2}{r}\lambda^2 + 1}$ (5-8)

I.1. Optimum size of the experiment

To determine the optimum size of the experiment we define the function

$$G = a\sum_{i=1}^{k} u_i P_i - bn \quad ; \text{ 'a' and 'b'} > 0$$

G may be called the gain function; u_i is the mean of the variety V_i; P_i is the probability that V_i will be selected;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'a' is a positive constant which multiplied by u_i gives the possible gain or worth of V_i. Possible gain multiplied by probability will give the expected worth of V_i and $a \sum_{i=1}^{k} u_i P_i$ is the expected gain or worth of the varieties being experimented on. From the possible gain cost of the experiment has to be subtracted, n denotes the size (replications) of the experiment and b is the cost per unit (replication) of experiment. Thus G gives the net gain or net worth of the varieties.

That size of the experiment will be considered optimum which maximizes this net gain. Instead of working with G we work with

$$G' = \frac{G}{a\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} = \frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} \Sigma u_i P_i - \left(\frac{b}{\sigma a}\right)^{2/3} n \qquad (5\text{-}9)$$

Maximization of G' is equivalent to maximization of G. To find $\Sigma \frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} u_i P_i$ we note that u_i has a regression on $\bar{x}_i$ and using the result (5-7) the regression of u_i on $\bar{x}_i$ is obtained to be $\frac{1}{\lambda^2\sigma^2/n + 1}$. Further the variety selected is the one which gives the max. $\bar{x}$. Therefore the value of $\frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} \Sigma u_i P_i$ will be given by the product of expected value of max $\frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} \bar{x}_i$ and the regression of u_i on $\bar{x}_i$. Further we note that $\bar{x}_i'$ ($\bar{x}_i' = \frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} \bar{x}_i$) has mean $\frac{1}{\sigma^{2/3}} \left(\frac{a}{b}\right)^{1/3} u_i$ and variance $\frac{\sigma^{2/3}}{n} \left(\frac{a}{b}\right)^{2/3}$; u_i is exponential

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with parameter λ and that will give $\frac{1}{\sigma^{2/3}}\left(\frac{a}{b}\right)^{1/3} u_i$ as exponential with parameter $\frac{\lambda\sigma^{2/3}}{c^{1/3}} = \lambda'$ $\left(c=\frac{a}{b}\right)$. Using the result (5-4) with the new exponential parameter and the variance we get

$$E(\max \bar{x}_i') = \frac{1}{\lambda'} + \frac{c^{1/3}\sigma^{1/3}}{\sqrt{(n)\pi}} + I(\sqrt{2}\lambda'\sigma^{1/3}c^{1/3}/\sqrt{n})\ \frac{e^{\lambda'^2 c^{2/3}(\sigma^{2/3}/n)}}{\lambda'} \tag{5-10}$$

Substituting (5-10) in (5-9) and using the regression of u_i on x_i we get

$$G' = \frac{1}{\frac{\lambda^2\sigma^2}{n}+1}\left[\frac{1}{\lambda'} + \frac{\sigma^{1/3}c^{1/3}/\sqrt{n}}{\sqrt{\pi}} + \frac{e^{\lambda'^2\sigma^{2/3}\sigma^{2/3}}}{\lambda'} I(\sqrt{2}\lambda'(\sigma c)^{1/3}/\sqrt{n})\right]$$

$-\left(\frac{1}{\sigma c}\right)^{2/3} n.$

and substitute $\left(\frac{1}{\sigma c}\right)^{2/3} n = n'$ and we will get

$$G' = \frac{1}{\frac{\lambda'^2}{n'}+1}\left[\frac{1}{\lambda'} + \frac{1}{\sqrt{(\pi n')}} + \frac{e^{\lambda'^2/n}}{\lambda'} I(\sqrt{2}\,\lambda'/\sqrt{n'})\right] - n' \tag{5-11}$$

For given values of λ' (5-11) depends upon on n' only. To find n' which maximizes (5-11), n' may be treated as continuous, differentiate with respect to n' and equate the differential to zero. The following equation will give the optimum n' for given value of λ'. From n' original n can be obtained by going backward through all the transformations. Differentiating G' gives the equation

$$\frac{\lambda'^2}{(\lambda'^2+n')^2}\left[\frac{1}{\lambda'} + \frac{1}{\sqrt{(\pi n)}} + \frac{e^{\lambda'^2/n'}}{\lambda'} I(\sqrt{2}\lambda'/\sqrt{n})\right] - \frac{e^{\lambda'^2/n'}}{(\lambda'^2+n')n'} I(\sqrt{2}\lambda'/\sqrt{n'}) - 1 = 0 \tag{5-12}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This involves only univariate normal integrals and tables for that are available. Graph 4 shows the relation of λ' with optimum n'. It is also to be noted that the value of n' determined from (5-12) really gives the maximum for G' and not minimum. In calculating n' from the above equation it was observed that it is a decreasing function of n', that will ensure that the second derivative will be negative. Interesting feature of (5-12) is that for $\lambda' = \infty$ or zero there does not exist a solution for it. If $\lambda' = \frac{\lambda\sigma^{2/3}}{(a/b)^{1/3}} = \infty$ that will imply one of the following possibilities

(i) λ infinity that will give u_i equal to zero.

(ii) 'a' zero implies there is no worth of the varieties.

(iii) 'b' infinity implies that the cost of the experiment is prohibitive.

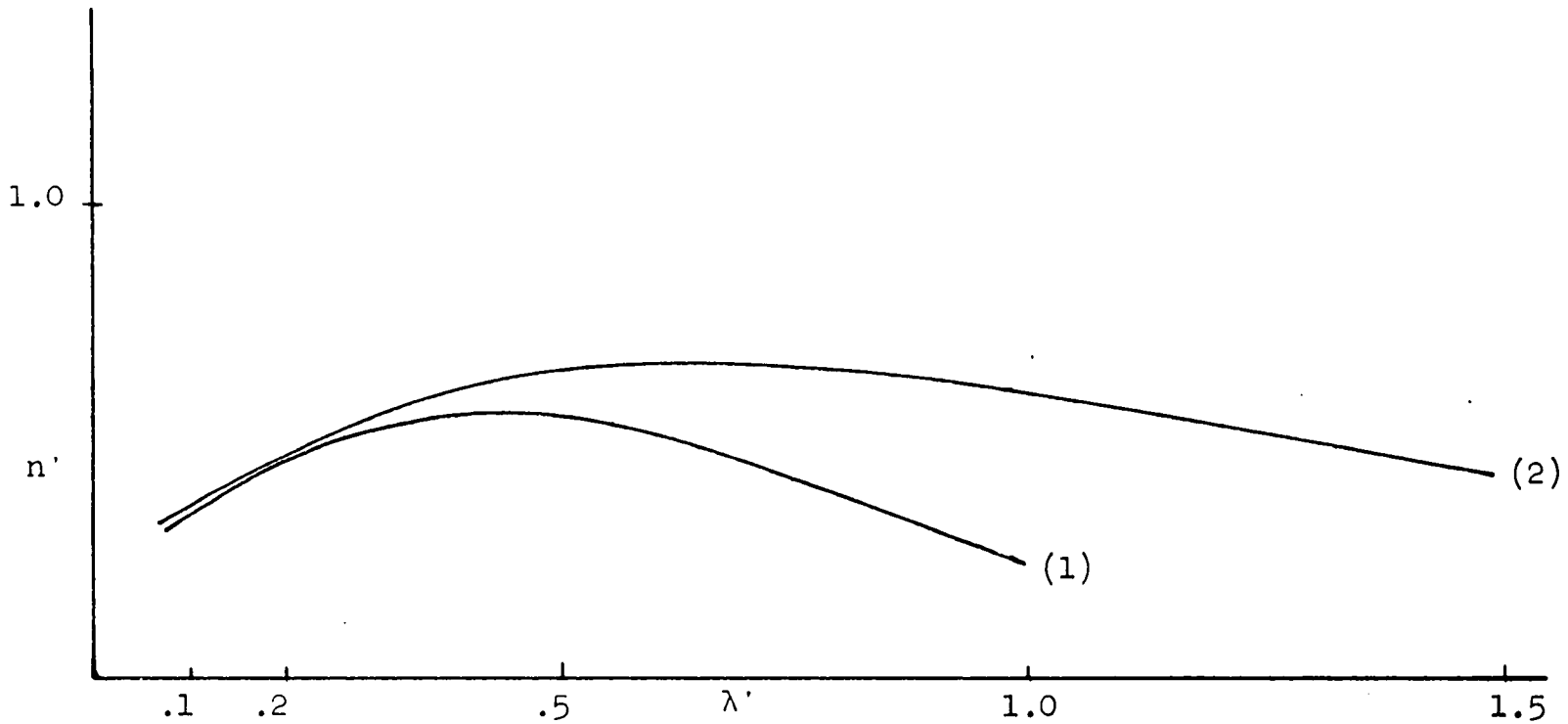
(iv) σ infinity will give the regression of u_i on $\bar{x}_i$ zero.

λ' is zero is equivalent to, 'b', the cost of the experiment is zero. Similar situation has been observed by Dunnett (19) that it does not pay to select from two populations. Finney (22) has observed that if the selection is from a large number of varieties it does not pay to select more than .27 fraction. Going through the same sort of arguments as before and using the results (5-6) and (5-7) for r equal to 2 we get the gain function G' for gamma with r equal to 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4



Optimum size of experiment for k=2 Optimum n' for given λ'

(1) Exponential (2) Gamma with r = 2

$$G' = \frac{2n'}{\lambda'^2+2n'}\left[\frac{2}{\lambda'} + \frac{3}{2\sqrt{\pi n'}} + \frac{3}{2}\frac{e^{\lambda'^2/n'}}{\lambda'} I(\sqrt{2}\lambda'/\sqrt{n'}) - \frac{\lambda' e^{\lambda'^2/n'}}{2n'} I(\sqrt{2}\cdot\lambda'/\sqrt{n'})\right] - n' \tag{5-13}$$

Optimum n' will be given by equating the differential (with respect to n') to zero, i.e.

$$\frac{2}{(\lambda'^2+2n')}\left[e^{\lambda'^2/n'} I(\sqrt{2}\lambda'/\sqrt{n'})\left(\frac{\lambda'^3}{n'^2} - \frac{\lambda'}{2n'}\right) - \frac{\lambda'^2}{2\sqrt{\pi}\cdot n'^{3/2}}\right] + \tag{5-14}$$

$$\frac{2\lambda'^2}{(\lambda'^2+2n')^2}\left[\frac{2}{\lambda'} + \frac{3}{2\sqrt{(\pi n')}} + e^{\lambda'^2/n'} I(\sqrt{2}\cdot\lambda'/\sqrt{n'})\left(\frac{3}{2\lambda'} - \frac{\lambda'}{n'}\right)\right] - 1 = 0$$

Again for λ' equal to zero or infinity there does not exist a solution for (5-14). If λ' is equal to zero all the expressions reduce to zero and we are left with (-1) equal to zero. If λ' is infinity second part is easily seen to be zero. The first part also goes to zero; that can be seen by noting that the first expression approaches the third when λ' approaches infinity. Thus again we are left with (-1) equal to zero. Figure 4 gives the optimum n' corresponding to λ' for the exponential and the gamma distribution with r equal to 2.

II. A Variation of the probability of correct selection problem

In the previous sections the form in which the selection problem has been posed is appropriate in a situation when the data is available on all the varieties at the same time. But

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in certain circumstances this may not be very appropriate. It is possible that the experimenter has finished his experiment, on the basis of that he has selected, possibly, the best variety; then comes another variety to be compared with the already selected variety. This type of problem is particularly likely to occur in pharmaceutical industry where any drug can be tested at any time and the time required for testing is not very long. The previous formulations are more applicable to agricultural experiments where an experiment requires at least one season and comparison among varieties grown over seasons will not be very appropriate due to seasonal disturbances.

The problem can be specified in this way. An experiment is conducted and the best variety has been selected out of k varieties. Then we have another variety, call it (k+1)th. We require the probability of correct selection between (k+1)th and the already selected one. The varietal means follow independently exponential distribution with same parameter λ; λ may be assumed different for all varieties, that will not alter the problem, only algebra will be more lengthy. This may be called kth stage selection, (in this particular sense).

So we need to determine

$$\text{Prob}(\text{correct selection between } (k+1)\text{th and the kth} \mid \text{kth is the best of the first k varieties})$$

$$= \text{Prob}(u_k > u_{k+1} \text{ and } v_k \text{ is selected} \mid u_k \text{ is the best of the first k})$$

$$+ \text{Prob}(u_{k+1} > u_k \text{ and } v_{k+1} \text{ is selected} \mid u_k \text{ is the best of the first k})$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

II.1. $\text{Prob}(V_k \text{ is selected}) = \text{Prob}(\bar{x}_k - \bar{x}_{k+1} > 0)$

$$= I(-y_k/E) \qquad (5\text{-}15)$$

$$y_k = u_k - u_{k+1}$$

u_k is the largest of the first k means and each follows independently exponential distribution with parameter λ, that will give

$$f(u_k\text{-max}) = \sum_{k=1}^{k} e^{-\lambda u_k} \lambda^k \int_0^{u_k} e^{-\lambda u_1} \cdots \int_0^{u_k} e^{-\lambda u_{k-1}} \prod_{j=1}^{k-1} du_j$$

$$= k\lambda e^{-\lambda u_k}(1-e^{-\lambda u_k})^{k-1} \qquad (5\text{-}16)$$

u_{k+1} also follows exponential distribution with parameter λ from that

$$f(u_k - u_{k+1} = y_k \mid u_{k+1}) = k\lambda e^{-\lambda y_k - \lambda u_{k+1}}(1-e^{-\lambda y_k - \lambda u_{k+1}}) \qquad y_k > 0$$

Multiplying the right hand side by the density function of u_{k+1} and integrating over the range zero to infinity for u_{k+1} will give the marginal density function of y_k

$$f(y_k) = \int_0^{\infty} k\lambda^2 e^{-\lambda y_k - 2\lambda u_{k+1}}(1-e^{-\lambda y_k - \lambda u_{k+1}})^{k-1} du_{k+1} \qquad (5\text{-}17)$$

$$= f_k$$

Integrating (5-15) with respect to (5-17) over the range zero to infinity will give the joint probability that u_k (already selected variety) is better than the newcomer and it is selected

$$\int_0^{\infty} f_k F(-y_k/E)\, dy_k \qquad (5\text{-}18)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Prob (V_{k+1} is selected) = Prob ($\bar{x}_{k+1} - \bar{x}_k > 0$)

$$= I\,(-y_{k+1}/E) \qquad y_{k+1} = u_{k+1} - u_k \tag{5-19}$$

$$f(u_{k+1} - u_k = y_{k+1} \mid u_k) = \bar{e}^{\lambda y_{k+1} - \lambda u_k} \quad \text{for } y_{k+1} > 0$$

Following similar arguments as before we will get

$$f(y_{k+1}) = \int_0^\infty k\lambda^2 e^{-\lambda y_{k+1} - 2\lambda u_k}(1-e^{-\lambda u_k})^{k-1}\, du_k = f_{k+1} \tag{5-20}$$

Integrating (5-19) with respect to (5-20) over the range zero to infinity will give the joint probability that V_{k+1} (the new variety) is better than the first selected and it is selected.

$$\int_0^\infty f_{k+1} F(-y_{k+1}/E)\, dy_{k+1} \tag{5-21}$$

Sum of (5-20) and (5-21) will give the probability of correct selection between the kth and the (k+1)th variety

$$\int_0^\infty f_k F(-y_k/E)\, dy_k + \int_0^\infty f_{k+1} F(-y_{k+1}/E)\, dy_{k+1} \tag{5-22}$$

Evaluation of (5-22) is pretty straightforward; it involves only univariate integral. The general term to be calculated is

$$\frac{\lambda}{\sqrt{(2\pi)}} \int_0^\infty \bar{e}^{n\lambda y} \int_{-y/E}^\infty \bar{e}^{t^2/2}\, dt\, dy = \frac{1}{n}\left[\tfrac{1}{2} + e^{\frac{n^2E^2\lambda^2}{2}}\, I(n\lambda E)\right]$$

As an illustration we work out (5-22) for k equal to 2

$$\text{for } k=2 \quad f_k = 2\lambda\left(\frac{e^{-\lambda y_k}}{2} - \frac{e^{-2\lambda y_k}}{3}\right) \tag{5-23}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and $f_{k+1} = \frac{1}{3} \times e^{-\lambda y_{k+1}}$ (5-24)

That will give the probability of correct selection

$$\int_0^\infty 2\lambda\left(\frac{e^{-\lambda y_k}}{2} - \frac{e^{-2\lambda y_k}}{3}\right) I(-y_k/E)\,dy_k + \frac{1}{3}\int_0^\infty e^{-\lambda y_{k+1}} I(-y_{k+1}/E)\,dy_{k+1}$$

$$= \left[\tfrac{1}{2} + e^{\lambda^2E^2/2} I(\lambda E)\right] - \frac{1}{3}\left[\tfrac{1}{2} + e^{2\lambda^2E^2} I(2\lambda E)\right] + \frac{1}{3}\left[\tfrac{1}{2} + e^{\lambda^2E^2/2} I(\lambda E)\right] \quad (5\text{-}25)$$

As n approaches zero (2-26) approaches $\frac{1}{2}$, i.e. random selection between the two varieties. As n approaches infinity (5-45) approaches unity. For k equal to 3 the corresponding expression is

$$3\left[\tfrac{1}{2}\left(\tfrac{1}{2} + e^{\lambda^2E^2/2} I(\lambda E)\right) - \tfrac{1}{6}\left(\tfrac{1}{2} + e^{2\lambda^2E^2} I(2\lambda E)\right) + \tfrac{1}{12}\left(\tfrac{1}{2} + e^{9\lambda^2E^2/2} I(3\lambda E)\right)\right.$$

$$\left. - \tfrac{1}{4}\left(\tfrac{1}{2} + e^{\lambda^2E^2/2} \cdot I(\lambda E)\right)\right] \quad (5\text{-}26)$$

Here again the expression approaches $\frac{1}{2}$ and unity as n approaches zero and infinity respectively.

III. Selection of t best varieties

Undoubtedly the experimenter will be interested in selecting the best variety but to safeguard himself against unforeseen circumstances or he may have reason to believe that later on funds will be available, thus enabling him to increase the scope of the experiment, he may like to select more than one, say 't' best varieties. We consider the case when the selection of the two best varieties is desired.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Firstly, the derivation will be for the case when the varietal means follow normal distribution with common known variance σ_o^2 and then it will be extended to exponential. To keep the things simple/U's, the means of u's in the case of normal are supposed to be equal and for the exponential case u's are supposed to have common parameter λ.

Let us denote the varieties by V_i and V_j, which are selected, that would imply

$$\bar{x}_m - \bar{x}_i < 0 \Rightarrow z_{mi} = \frac{\bar{x}_m - \bar{x}_i - (u_m - u_i)}{E} < \frac{-(u_m - u_i)}{E} \quad \text{for } m \neq j \neq i$$

$$\text{and } \bar{x}_m - \bar{x}_j < 0 \qquad z_{mj} = \frac{\bar{x}_m - \bar{x}_j - (u_m - u_j)}{E} < \frac{(u_m - u_j)}{E} \quad m \neq j \neq i$$

$$\bar{x}_k - \bar{x}_i < 0 \qquad z_{ki} = \frac{\bar{x}_k - \bar{x}_i - (u_k - u_i)}{E} < \frac{(u_k - u_i)}{E}$$

$$\bar{x}_k - \bar{x}_j < 0 \qquad z_{kj} = \frac{\bar{x}_k - \bar{x}_j - (u_k - u_j)}{E} < \frac{-(u_k - u_j)}{E} \quad k \neq j \neq i$$

z's are standardized normal variables with correlation structure

$$\rho(z_{mi}\ z_{kj}) = \tfrac{1}{2} \quad i \neq j \quad m = k$$

$$= \tfrac{1}{2} \quad i = j \quad m \neq k$$

$$0 \quad i \neq j \quad m \neq k$$

$$1 \quad i = j \quad m = k$$

From this we will get the probability that V_i and V_j are selected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$F_{2(k-2)\ \frac{1}{2},\frac{1}{2},0}\left[\frac{-(u_k-u_i)}{E},\frac{-(u_k-u_j)}{E},\frac{-(u_m-u_i)}{E},\frac{-(u_m-u_i)}{E}\right] k \text{ and } m \neq j \neq i$$

$$=F_{2(k-2),\frac{1}{2},\frac{1}{2},0}\left[-y_{ki}/E,-y_{kj}/E;-y_{mi}/E,-y_{mj}/E\right] \qquad (5\text{-}27)$$

The probability depends upon u's which are supposed to follow independently normal distribution. Thus $u_m-u_i=y_{mi}$ and $u_m-u_j=y_{mj}$ also are normal with zero means and variance $2\sigma_o^2$. These y's are also correlated and their correlation structure is similar to z's. Let us denote the normal density function of y's by $f_{y_{ij}}$. Then integrating (5-27) with respect to $f_{y_{ij}}$ over the negative values of y's will give the probability that V_i and V_j are the two best varieties and they are selected

$$P_{ij}=\int_{-\infty}^{0}\int_{-\infty}^{0} f_{y_{ij}} F_{2(k-2),\frac{1}{2},\frac{1}{2}\ 0}\left(\qquad\qquad\right)\prod_{\substack{m\neq i\\ \neq j}} dy_{mi}\,d_{mj} \qquad (5\text{-}28)$$

But any two of the k varieties could be the best and selected. So the probability of correct selection will be given by summing over all pairs i and j. However, when all U's are equal all the terms will have identical values and to get the probability value we need to multiply (5-28) by

$$\binom{k}{2}=\frac{k!}{2!\ (k-2)!}$$

The method can be extended for t greater than 2 in a straightforward manner. But to get the numerical values is altogether another matter. Even for the simplest case of k equal to three out of which we want to select two

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

varieties, (5-28) will be a four variate normal integral. So far we do not know how to evaluate such integrals except for certain special correlation structures. Nevertheless for the particular case of k equal to three and 't' equal to two (5-28) can be evaluated by changing it to a different but equivalent form. The probability that V_1 and V_2 are better than V_1 and they appear as such in the experiment is the same as that V_3 is inferior to V_1 and V_2 and appears as such in the experiment. In this form the problem is a special case of the general problem that V_i is inferior to all others and in the experiment it appears as such. So this general problem is dealt with,

V_i is rejected over all others $\Rightarrow \bar{x}_j - \bar{x}_i > 0$ for $j \neq i$

$$z_j = \frac{\bar{x}_j - \bar{x}_i - (u_j - u_i)}{E} \geq \frac{-(u_j - u_i)}{E}$$

z's are standardized normal variables and they are correlated

$$\rho(z_i z_j) = \tfrac{1}{2} \quad j \neq i$$

$$= 1 \quad j = i$$

That will give Probability that V_i is rejected over all other varieties

$$I_{k-1,\frac{1}{2}} \left[\frac{-(u_j - u_i)}{E} \; ; \quad j \neq i \right] \qquad (5\text{-}29)$$

And u's follow normal distribution ; from that we get the probability that u_i is the least of all other u's

$$I_{k-1,\frac{1}{2}} \left[-(U_j - U_i)/\sqrt{2}\sigma_o \; ; \; j \neq i \right] \qquad (5\text{-}30)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Integrating (5-29) with respect to (5-30) will give the probability that u_i is the least of all other u's and V_i will be rejected over all varieties.

$$P_i = \int_{b_1}^{\infty} \cdots \int_{b_k}^{\infty} f(y_j;\ j\neq i)\ I_{k-1,\frac{1}{2}}(-y_j/E;\ j\neq i) \prod_{j\neq i} dy_j \qquad (5\text{-}31)$$

$$b_j = -(U_j - U_i)/\sqrt{2}\,\sigma_o$$

$$y_j = (u_j - u_i)$$

and $f(y_j;\ j\neq i)$ is a k-1 variate normal density function with all correlations equal to $\frac{1}{2}$.

(5-29) is equivalent to $\text{Prob}(x_j > -y'_j T\ ;\ j\neq i)$, $T = \sqrt{2}\sigma_o/E$ $y'_j = (u_j - u_i)/\sqrt{2}\sigma_o$

Substitute $z_j = (x_j - Ty'_j)/\sqrt{(T^2+1)}$: x_j, y'_j, z_j are standardized normal variates, y_j and z_j are independent of each other.

$$\text{Cov}(x_s x_t) = \text{Cov}(y'_s y'_t) = \text{Cov}(z_s z_t) = \tfrac{1}{2}$$

$$\text{Cov}(z_s y'_t) = \sqrt{T^2}/\sqrt{(T^2+1)} = \rho \quad \text{for } s = t$$

$$= \tfrac{1}{2}\rho \quad \text{for } s \neq t$$

$$\text{Cov}(x_j z_i) = 0 \text{ for all } i \text{ and } j$$

With this transformation (5-31) reduces to

$$P_i = \int_{b_j}^{\infty} \cdots \int_{0}^{\infty} f(y'_j;\ z_j;\ j\neq i) \prod_{j\neq i}^{k} dy_j \prod_{j\neq i}^{k} dz_j : f(y'_j;\ z_j;\ j\neq i)$$ is

2(k-1) variate normal density function with the correlation structure

y's	z's
A	B
B	A

$a_{ij} = \tfrac{1}{2}$ $i\neq j$; $= 1$ $i=j$

$b_{ij} = \rho\,\tfrac{1}{2}$ $i\neq j$; $= \rho$ $i=j$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The range of integration for the z's is from zero to infinity and for the y's from $-(U_j-U_i)/\sqrt{2}\sigma_o$ to infinity. However if all the U's are equal the range of integration for all the variates will be from zero to infinity and the probability that the poorest variety is picked up as such is

$$P = \Sigma P_i = kP_i = k \int_0^\infty \int_0^\infty f(y'_j,\ z_j;\ j\neq i) \prod_{j\neq i}^{k} dy'_j \prod_{j\neq i}^{k} dz_j \qquad (5\text{-}32)$$

III.1. If the varietal means are supposed to follow independently exponential distribution then the probability that V_i and V_j are the best and they are selected will be given by integrating (5-27) with respect to f (f to be defined below) over negative values of y's.

$$u_m-u_i = y_{mi} \qquad u_m-u_j = y_{mj}$$

and $$u_k-u_i = y_{ki} \qquad u_k-u_j = y_{kj}$$

m and k $\neq$ j $\neq$ i

f is the density function of y's as defined above. For the time being it has not been possible to have the joint density function (f) for the y's for general k. However, the simplest case of k equal to 3 and t equal to 2 can be dealt with in a straightforward manner; it is a special case of the problem that the poorest variety is picked up as such.

$$\text{Prob}(V_i \text{ is picked up as the poorest variety}) = I_{k-1,\frac{1}{2}}(-y_j/E\ ;\ j\neq i) \qquad (5\text{-}33)$$

and for the probability that u_i is the least we need the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

density function of y_j's $(u_j - u_i = y_j)$ for positive y_j. From (2-20) of chapter 2 we can have that. Integrating (5-33) with respect to (2-20) will give the probability that u_i is the least of all and V_i is picked up as such. For all λ's equal it is

$$P_i = \frac{(k-1)!}{k} \int_0^\infty \int_0^{y_1} \cdots \int_0^{y_{k-1}} \lambda^{k-1} e^{-\sum_{j\neq i} \lambda y_j} I_{2,\frac{1}{2}}(-y_j/E \; ; \; j\neq i)\, dy_k \cdots dy_{i+1} \cdots dy_{i-1} \cdot dy_1 \tag{5-34}$$

IV. Complete Ranking

Still another variation of the selection problem can be the complete ranking, i.e., the experimenter requires the varieties to be arranged in order of magnitude, $V_1 < V_2 < \ldots < V_k$ For given n (sample size) the probability that the varieties are ranked correctly can be determined as follows. Let us take a particular ranking say, R_1 ; $V_1 < V_2 < \ldots < V_k$.

The numbering of the varieties is immaterial because they can be renumbered to have a particular permutation. Ranking 'R_1' would mean that the sample means appear as

$\bar{x}_1 < \bar{x}_2 < \ldots < \bar{x}_k$ that would imply that

$z_1 = \bar{x}_1 - \bar{x}_2 < 0$

$z_2 = \bar{x}_2 - \bar{x}_3 < 0$

$\ldots\ldots\ldots\ldots\ldots$

$z_{k-1} = \bar{x}_{k-1} - \bar{x}_k < 0$

The z's are normal with means $(u_j - u_{j+1},\ j=1,\ \ k-1)$ with equal variances $E^2 = 2\sigma^2/n$ and they are correlated

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\rho(z_i z_j) = 1 \quad i=j$$
$$= -\tfrac{1}{2} \quad i=j \pm 1$$
$$= 0 \quad j-1<i<j+1$$

That will give the probability that we observe the above mentioned ranking

$$F_{k-1,(1,-1/2,0)}\left[-(u_\rho - u_2)/E, \ldots, -(u_{k-1}-u_k)/E \right]$$

$$F_{k-1,(1,-1/2,0)} \quad (-y_j/E \quad j\neq k) \qquad y_j = u_j - u_{j+1} \qquad (5\text{-}35)$$

Under the assumption that u's follow normal distribution with common known variance the y's will be normal with the same correlation structure as z's and variance $2\sigma_o^2$. Probability that the varietal means are also ranked as R_1 will be given by

$$F_{k-1,(1,-1/2,0)}\left[-(U_j - U_{j+1})/\sqrt{2}\sigma_o \; ; \; j=i,\ldots k-1) \right] \qquad (5\text{-}36)$$

So we get the probability that the varietal means are in R_1 ranking and their sample means also appear in the same ranking by integrating (5-35) with respect to (5-36)

$$\text{Prob}(R_1) = \int_{-\infty}^{-(U_1-U_2)/\sqrt{2}\sigma_o} \int_{-\infty}^{-(u_{k-1}-u_k)/\sqrt{2}\sigma_o} f(y_j;\ j=1,\ldots k-1) F(-y_j\ E;\ j=1,\ldots,k-1) \prod_{j=1}^{k-1} dy_j \qquad (5\text{-}37)$$

If all u's are equal the integration will be from zero to minus infinity. The varietal means can have k! rankings and correspondingly k! correct rankings. Therefore to get the probability of correct ranking we have to sum over all possible permutations; in the case all U's are equal it will be multiplication of (5-37) by k!. Again for numerical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

results we are very much handicapped. Even for the simplest case of k equal to 3 (5-37) will be a four variate normal integral.

IV.1. In the case of exponential distribution we can do little better but, not much. Probability that the u's are in R_i ranking (as defined for the normal in IV) is given by

$$\prod_{i=1}^{k} \lambda_i \int_0^{\infty} e^{-\lambda u_k} \int_0^{u_{k-1}} \cdots \int_0^{u_2} e^{-\lambda_1 u_1} \, du_1 \ldots du_k \qquad (5\text{-}38)$$

Thus in the exponential case the probability that the u's are in R_1 ranking and we get their sample means in the same ranking is given by integrating (5-35) with respect to (5-38)

$$\text{Prob}(R_1) = \prod_{i=1}^{k} \lambda_i \int_0^{\infty} e^{-\lambda_k u_k} \int_0^{u_{k-1}} \cdots \int_0^{u_2} e^{-\lambda_1 u_1} F_{k-1,(1,-\frac{1}{2},0)}(-y_j/E;) \prod_{j=1}^{k} du_j \qquad (5\text{-}39)$$

This will require the evaluation of multiple integrals (over various regions) but, for k equal to 3 we can manipulate the integrals in such a way that the numerical results can be obtained with the help of existing tables.

$$\text{For } k = 3,\ P(V_1 < V_2 < V_3) = F_{2,-1/2}\left[-(u_1-u_2)/E, -(u_2-u_3)/E \right]$$

$$= F_{2,-1/2}\left[(u_2-u_1)/E, (u_3-u_2)/E \right] \qquad (5\text{-}40)$$

We work the joint density function for $u_2-u_1=y_1$ and $u_3-u_2=y_2$ (y_1 and $y_2 > 0$)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$f(u_3-u_2=y_2 \mid u_2) = \lambda_3 e^{-3y_2 - \lambda_3 u_2} \qquad y_2 > 0$$

$$f(u_2-u_1=y_1 \mid u_2) = \lambda_1 e^{\lambda_1 y_1 - \lambda_1 u_2} \qquad u_2 > y_1 \text{ and } y_1 > 0$$

for given u_2, f and f' are independent, thus giving

$$f(y_1\ y_2 \mid u_2) = \lambda_1 . \lambda_3\ e^{-\lambda_2 y_2}\ e^{+\lambda_1 y_1}\ e^{-u_2(\lambda_1+\lambda_3)} \qquad (5\text{-}41)$$

Multiplying the right hand side by the density function of u_2 and integrating out u_2 will give

$$f(y_1 y_2) = \int_{y_1}^{\infty} \lambda_1\lambda_2\lambda_3 . e^{-\lambda_2 y_2}\ e^{-\lambda y_1} . e^{-u_2(\lambda_1+\lambda_2+\lambda_3)}$$

$$= \frac{\lambda_1\lambda_2\lambda_3\ e^{-\lambda_2 y_2 - y_1(\lambda_1+\lambda_2)}}{\lambda_1+\lambda_2+\lambda_3} \qquad (5\text{-}42)$$

Assuming $\lambda_1=\lambda_2=\lambda_3=\lambda$ (5-42) reduces to $\frac{\lambda^2}{3} . e^{-\lambda_2 y_2 - 2\lambda_1 y_1}$ (5-43)

Integrating (5-40) with respect to (5-43) gives the joint probability that the means of the three varieties are in the order $(u_1 < u_2 < u_3)$ and they are selected as such

$$\frac{\lambda^2}{3} \int_0^{\infty} \int_0^{\infty} e^{-\lambda y_2 - 2\lambda y_1} F_{2,-\frac{1}{2}}(y_1/E,\ y_2/E)\, dy_1 dy_2 \qquad (5\text{-}44)$$

Summing over all possible permutations of V_1, V_2 and V_3 will give the probability of ranking the varieties in their correct order. When all λ's are equal summation will be replaced by multiplication by 3!. Evaluation of (5-44) is pretty straightforward. First integrate with respect to y_1 and then with respect to y_2. Going through simple algebra gives the result,

$$(5\text{-}44) = \frac{1}{6}\Big[F_{2,-\frac{1}{2}}(0,0) + \frac{\lambda^2E^2}{2\pi} \int_{E\lambda}^{\infty} \int_{-\infty}^{\infty} f\ dx\ dy + e^{-\lambda E} \frac{2\lambda^2E^2}{2\pi} \int_{-\infty}^{2E\lambda} \int_{-\infty}^{-\lambda E} f\ dx dy + e^{3\lambda^2E^2/2} \int_{\frac{3}{2}\lambda E}^{\infty} \int_{0}^{\infty} f\, dx dy$$

f is bivariate density function with correlation $(-\frac{1}{2})$ (5-45)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From (5-45) it is clear that as λE approaches infinity i.e. when n approaches zero the last three terms approach zero and thus it reduces to $\frac{1}{6}F_{2,-\frac{1}{2}}(0,0)$ and that will give the probability of correct ranking equal to 1/6. That is exactly the probability of correct ranking without an experiment. On the other hand if λE approaches zero i.e. n approaches infinity the last three terms of (5-45) represent the fourth, second and first quandrant respectively of the bivariate distribution and $F_{2,-\frac{1}{2}}(0,0)$ represents the third quadrant and thus the sum of the four is one. Hence the probability of correct ranking is unity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 6

SELECTION ON TWO CHARACTERS

I.

So far we have considered various problems concerned with the selection of the best variety out of k varieties. That variety was taken to be the best which gave the maximum sample mean for the character under observation. In some cases it may be desirable to base selection on more than one character. As we go beyond one character it becomes very difficult to define what may be called the best variety. Obviously, if one can find a variety exceeding others in all the characters specified, he should select that variety but, that will be in most of the cases asking for too much because none of the varieties may be exceeding the others in all the characters. Very little has been done for the selection problems when more than one character is involved. A very simple problem involving two characters is discussed in the following sections.

Selection is to be based on two characters, and the decision to select or reject depends upon the sample means for the two characters. The characters X_1 and X_2 may be correlated but as the problem is formulated the correlation does not come into picture. To give a precise formulation the problem may be stated as thus. First selection is based on character X_1, out of k varieties 't' best are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selected and out of those t the variety which exceeds in χ_2 will be finally selected. In the particular case when t = 1 it reduces to selection on X_1 alone. On the other extreme when 't' is equal to k it reduces to selection on X_2 alone. Thus the selection on a single character may be viewed as particular cases of selection on two characters. The selection problem on two characters can be solved either taking some relevant configuration or a priori distribution of the means of the varieties for the two characters.

II.

First the solution is given using a particular configuration.

x_{1ij} is the observation for character χ_1 on variety V_i, normal with mean u_{1i} and common known variance σ_1^2. Similarly x_{2ij} is the observation for character χ_2 on variety V_i, normal with mean u_{2i} and common known variance σ_2^2;(j=1, ... n.)

The following configuration is assumed for the varietal means. The varieties in the upper group (consisting of t varieties) are at least δ superior to the lower group in χ_1 and there is one variety V_i in the upper group which is at least Δ superior to all others in χ_2. We may assume that the varieties in the lower group (with respect to χ_1) are also Δ inferior to V_i in χ_2. (In some cases it may be more realistic to assume that they are not superior to V_i by more than a certain amount Δ_1) and the varieties in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

upper group are equivalent to V_i with respect to χ_1. In some cases it may be more realistic to assume that the varieties in the upper group are at most δ_1 superior to V_i in χ_1. These changes can be accommodated by obvious modifications of the procedure to be followed below. Probability of correct selection is calculated below. The selection will be correct if V_i as described above is selected.

Without loss of generality we can number the k varieties $V_1, V_2, \ldots, V_t, V_{t+1}, \ldots V_k$ and let $V_1, \ldots V_t$ be the group superior to $V_{t+1}, \ldots V_k$ in χ_1 and let V_1 be superior to all others in χ_2 by Δ i.e.

$u_{1t+1}=u_{1t+2}=u_{1t+3}, \ldots =u_{1k} < u_{11}-\delta$ and $u_{11}=u_{12}= \ldots =u_{1t}$

and $u_{21}-\Delta > u_{22}= \quad =u_{2k}$.

Then the probability of correct selection = P(On χ_1, V_1 is not rejected i.e. V_1 is one of the top t varieties in χ_1 and amongst those t it is the best in χ_2) and may be expressed as

$$\sum_{i=1}^{m} P(a_i)P(b_i \mid a_i) \qquad m = \binom{k-1}{t-1}$$

Where a_i is the event that V_i is not rejected on χ_1 and b_i is the event that it is selected on χ_2. The event that V_1 is not rejected on χ_1 is a compound event consisting of all those events (a_i's) in which any (t-1) varieties of the remaining k-1 are the topmost along with V_1. In all there will be $\binom{k-1}{t-1}$ a's. In the above problem the events are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$a_1 = V_1, V_2, \quad , V_t$ are selected on $\mathfrak{X}_1$

$a_2 = V_2, V_3, \quad V_t, V_{t+1}$ are selected on $\mathfrak{X}_1$

..... and so on

For each event a_i there is a certain probability $(b_i | a_i)$ - denoting the probability and the event by the same letter - that V_1 will be selected. To indicate the method we calculate the probability for a_1 and $(b_1 | a_1)$. The event

$$\begin{array}{llll} a_1 = V_1, V_2, & , V_t > V_{t+1} \cdots , V_k & & \text{i.e.} \\ V_1 > V_{t+1}, & , V_k & \bar{x}_{1t+1} - \bar{x}_{11} < 0, & , \bar{x}_{1k} - \bar{x}_{11} < 0 \\ V_2 > V_{t+1}, & , V_k & \bar{x}_{1t+1} - \bar{x}_{12} < 0, & , \bar{x}_{1k} - \bar{x}_{12} < 0 \\ \cdots\cdots\cdots\cdots & & \cdots\cdots\cdots & \cdots\cdots\cdots \\ V_t > V_{t+1}, & , V_k & \bar{x}_{1t+1} - \bar{x}_{1t} < 0, & , \bar{x}_{1k} - \bar{x}_{1t} < 0 \end{array} \qquad (6\text{-}1)$$

Define $\bar{x}_{1t+1} - \bar{x}_{11} = z_{1t+1}$ and $\bar{x}_{1t+1} - \bar{x}_{1t} = z_{tt+1}$,

With this notation (6-1) becomes

$$\begin{array}{ll} z_{1t+1} < 0, & z_{1k} < 0 \\ \cdots\cdots & \cdots\cdots\cdots \\ z_{tt+1} < 0 & z_{tk} < 0 \end{array} \qquad (6\text{-}2)$$

The z's are normal with means $-\delta$ and variances $2\sigma_1^2/n$ and correlation structure

$$\begin{aligned} \rho(z_{ij}, z_{pq}) &= \tfrac{1}{2} \quad \text{if} \quad i \neq p \quad j = q \\ &= \tfrac{1}{2} \qquad i = p \quad j \neq q \\ &= 0 \qquad i \neq p \quad j \neq q \\ &= 1 \qquad i = p \quad j = q \end{aligned}$$

From that we get the Prob (a_1) = Prob $(V_1, \quad , V_t > V_{t+1}, \quad , V_k)$

$$= \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{t(k-t)}{2}}} \int_{-\infty}^{\frac{\delta}{\sigma_1/n}} \cdots \int_{-\infty}^{\frac{\delta}{\sigma_1/n}} e^{-\frac{1}{2} z'_{ij} \Sigma^{-1} z_{ij}} dz_{1t+1} \cdots dz_{tk} \qquad (6\text{-}3)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Σ is the correlation matrix of z's and z'_{ij} and z_{ij} are the row and the column vectors of the variables which in conjunction with Σ determine the density function.

To calculate $(b_1|a_1)$ we note that V_1, V_2, ,V_t have been selected. Now that V_1 is selected implies $V_1 > V_2$, ,V_t in X_2.

$$\begin{array}{llll} \text{i.e. } \bar{x}_{22} - \bar{x}_{21} \leq 0 & \text{or} & y_{21} < 0 & \\ \bar{x}_{23} - \bar{x}_{21} \leq 0 & \text{or} & y_{31} < 0 & (6\text{-}4) \\ \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots & \\ x_{2k} - x_{21} \leq 0 & \text{or} & y_{k1} < 0 & \end{array}$$

y's are normal with means $(-\Delta)$ and variance $2\sigma_2^2/n$ and correlation structure

$$\rho\,(y_{i.1}\, y_{j.1}) = \tfrac{1}{2} \quad i \neq j$$
$$= 1 \quad i = j$$

Thus $(b_1|a_1)$ will be

$$F_{t-1,\frac{1}{2}}\left(\frac{\Delta}{\sigma_2\sqrt{\left(\frac{2}{n}\right)}}, \quad \frac{\Delta}{\sigma_2\sqrt{\left(\frac{2}{n}\right)}}\right) \qquad (6\text{-}5)$$

From (6-3) and (6-4) can be obtained the probability of $(a_1)(b_1|a_1)$

$$\text{Prob}(a_1)(b_1|a_1) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{t(k-t)}{2}}} \int_{-\infty}^{\frac{\delta}{\sigma\sqrt{2/n}}} \int_{-\infty}^{\frac{\delta}{\sigma\sqrt{2/n}}} \exp -\tfrac{1}{2}(z'_{ij}\Sigma^{-1}z_{ij})\,dz_{1t+1}\cdots dz_{tk}$$

$$\times F_{t-1,\frac{1}{2}}\left(\frac{\Delta}{\sigma_2\sqrt{(2/n)}}, \quad \frac{\Delta}{\sigma_2\sqrt{(2/n)}}\right) \qquad (6\text{-}6)$$

Similarly we can calculate the probabilities for a_2, $(b_2|a_2)$ and for other events. $a_2 = V_1, V_3, \quad V_t, V_{t+1} > V_2, \quad V_{t+2}, \quad V_k$. The procedure is just like the previous one but instead of all z's having mean $-\delta$ some will have zero mean e.g.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\bar{x}_{12}-\bar{x}_{11} = z_{12}$ has zero mean ; and some will have mean positive δ. Of course under the assumption that V_1 is Δ superior to all other varieties in X_2, b_j will be the same for every a_j. So the probability of correct selection can be written,

$$F_{t-1,\frac{1}{2}}\left(\frac{\Delta}{\sigma_2\sqrt{(2/n)}}, \quad , \frac{\Delta}{\sigma_2\sqrt{(2/n)}} \right) \cdot \sum_{i=1} (a_i) \qquad (6\text{-}7)$$

II.1.

For higher values of k and t we need the multiple integrals; and even then the method may be fairly laborious. However, for k equal to 3 and t equal to 2 the method can be used with advantage. That requires the evaluation of bivariate integrals only.

Let the three varieties be V_1, V_2 and V_3 and assume V_1 and V_2 are δ superior in X_1 to V_3 and V_1 is Δ superior to both V_3 and V_2 in X_2. The probability of correct selection i.e. the selection of V_1 is given by the following two combination of events.

a_1= on X_1, V_1 and V_2 are selected and then $(b_1|a_1)$, i.e V_1 is selected on X_2

a_2= on X_1, V_1 and V_3 are selected and then $(b_2|a_2)$, V_1 is selected on X_2

a_1 implies $z_{13} = \bar{x}_{13}-\bar{x}_{11} < 0$ and $z_{23} = \bar{x}_{13}-\bar{x}_{12} < 0$, z's have have means $(-\delta)$ variance $2\sigma_1^2/n$ and correlation $\frac{1}{2}$. That will give probability for a_1-

$$\text{Prob}(a_1) = F_{2,\frac{1}{2}}\left(\frac{\delta}{\sigma_1\sqrt{(2/n)}}, \frac{\delta}{\sigma_1\sqrt{(2/n)}}\right) \qquad (6\text{-}8)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the event $(b_1|a_1)$ implies $y_{21} = \bar{x}_{22} - \bar{x}_{21} < 0$; y_{21} is normal with mean $-\Delta$ and variance $2\sigma_2^2/n$ and that will give the probability of the event

$$\text{Prob}\ (b_1|a_1) = F\left(\frac{\Delta}{\sigma_2\sqrt{2/n}}\right) \tag{6-9}$$

Going through similar steps we will get $P(a_2) = F_{2,\frac{1}{2}}\left(\frac{-\delta}{\sigma_1\sqrt{(2/n)}}, 0\right)$ (6-10)

As before $P(b_2|a_2) = P(b_1|a_1)$. Hence the probability of correct selection

$$= F\left(\frac{\Delta}{\sigma_2\sqrt{(2/n)}}\right)\left\{F_{2,\frac{1}{2}}\left(\frac{\delta}{\sigma_1\sqrt{(2/n)}}, \frac{\delta}{\sigma_1\sqrt{(2/n)}}\right) + F_{2,\frac{1}{2}}\left(\frac{-\delta}{\sigma_1\sqrt{(2/n)}}, 0\right)\right\}$$

$$= A\ (B + C) \tag{6-11}$$

As n increases 'A' increases and 'B' increases but 'C' decreases. With n increasing to infinity 'A' and 'B' approach unity where as 'C' approaches zero. On the other hand if n approaches zero (6-11) approaches 1/3, i.e. random selection without any experiment.

III.

Alternative to any configuration of the means of the varieties in respect to two characters is the assumption that the means themselves follow some distribution. In this section general framework for the probability of correct selection (as defined in section II) is given when the varietiel means follow some distribution <u>a priori</u>; and complete solution for k equal to 3 and t equal to 2 is given. Varietal means are supposed to follow exponential distribution; u_{1i}'s, i.e., means with respect to X_1, all have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the same parameter λ_1 and u_{2i}'s, i.e., the means with respect to X_2 all have the same parameter λ_2.(λ_1's i.e. λ_{11}, λ_{12}, λ_{13} could be different but that does not effect the result apart from increasing the algebraic work, same holds for λ_2's. Probability of correct selection will be equivalent to this - a variety V_i has the greatest mean for X_2 amongst those 't' varieties which are the topmost in X_1 and it is finally selected. It may be noted that no assumption is made about the relation between the means for the character X_1 among the top 't' varieties themselves and likewise no relation is supposed between the means of the varieties in the lower group (for X_1) and the variety V_i to be selected, which is in the upper group (for X_1) for the character X_2.

Denote the set of 't' varieties which are the topmost in X_1 as T and the others k-t varieties by $\bar{T}$. Two sets, T_1 and T_2 are equal if and only if V_i in T_1 implies it is contained in T_2 also. Any 't' of the k varieties may have larger u_{1i} than the rest and any of the apparently superior 't' varieties may have larger u_2 for the character X_2. Denote different T's as T_1, T_2 .. T_m, $m = \binom{k}{t}$. Also denote by T_i the probability of correct selection when T_i is the group with 't' top varieties. Thus the probability of correct selection will be $\sum_{i=1}^{m} T_i$

When T_i is the group with larger u_{1i}'s in X_1 any of the 't' varieties can have larger u_2 in X_2. Again denote by t_{in} as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the probability that group T_i is with larger means in X_1 and variety V_n has the greatest u_2 in X_2 and is selected. Thus the probability of correct selection is given by

$$\sum_{i=1}^{m} \sum_{n=1}^{t} t_{in}$$

III.1

For higher values of k and t the number of integrals (which will be multiple integrals) increases pretty fast. Solution for the simplest case of k=3 and t=2 is discussed below.

Let the three varieties be V_1, V_2 and V_3. With t=2 there will be 3 T's, e.g., V_1V_2, V_2V_3 and V_1V_3. We solve for T_1 which consists of V_1 and V_2. So the problem is reduced to find the joint probability that V_1 and V_2 are superior to V_3 in X_1 and V_1 is superior to V_2 in X_2 and V_1 is finally selected. No relation is supposed between the means of V_1 and V_2 for X_1 and between the means of V_1 and V_3 for X_2.

For V_1 to be finally selected it is necessary that it is not rejected on X_1, that can happen in two ways, e.g., on X_1 either V_1 and V_2 are selected or V_1 and V_3 are selected and then V_1 is finally selected.

$$\text{Prob}(V_1 \text{ \& } V_2 \text{ are selected on } X_1) = P(\bar{x}_{13}-\bar{x}_{11}<0 \text{ and } \bar{x}_{13}-\bar{x}_{12}<0)$$

$$= F_{2,\frac{1}{2}}\left[\frac{(u_{11}-u_{12})}{E}, \frac{(u_{12}-u_{13})}{E}\right] \qquad (6\text{-}12)$$

u_{1i}'s are independently exponentially distributed with the same parameter λ_1. Joint density function of $y_1=u_{11}-u_{13}$ and $y_2=u_{12}-u_{13}$, for positive y's can be had from (2-20) of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

section III chapter 2 and may be written as

$$f(y_1y_2) = \frac{\lambda_1^2 e^{-\lambda_1 y_1 - \lambda_1 y_2}}{3} \qquad (6\text{-}13)$$

Then integrating (6-12) with respect to (6-13) over the positive range of y's will give the joint probability that V_1 and V_2 are in T_1 and they are selected on X_1.

$$\frac{2}{3} \int_0^{\infty} \int_0^{y_1} \lambda_1^2 \; e^{-\lambda_1 y_1 - \lambda_1 y_2} F\left(\frac{y_1}{E}, \frac{y_2}{E}\right) dy_2 dy_1 \qquad (6\text{-}14)$$

Given that V_1 and V_2 are selected on X_1, probability that V_1 is selected is

$$\text{Prob}\ (\bar{x}_{22} - \bar{x}_{21} < 0) = F\left(\frac{u_{21} - u_{22}}{E_2}\right) \qquad (6\text{-}15)$$

And density function of $y' = u_{21} - u_{22}$, for positive y'is given as

$$f(y') = \frac{\lambda_2}{2} e^{-\lambda_2 y'} \qquad (6\text{-}16)$$

Conditional probability that V_1 and V_2 are selected on X_1 and V_1 is superior to V_2 in X_2 and it is selected will be given by integrating (6-15) with respect to (6-16) and it is

$$\int_0^{\infty} \frac{\lambda_2}{2} e^{-\lambda_2 y'} F\left(\frac{y'}{E_2}\right) dy' \qquad (6\text{-}17)$$

Multiplying (6-14) and (6-17) will give the probability that V_1 and V_2 are better in X_1, they are selected on X_1 and V_1 is better than V_2 in X_2 and is finally selected

$$\frac{2}{3} \int_0^{\infty} \int_0^{y_2} \lambda_1^2 \; e^{-\lambda_1 y_1 - \lambda_1 y_2} F_{2,\frac{1}{2}}\left[\left(\frac{y_1}{E}, \frac{y_2}{E}\right)\right] dy_1 dy_2 \left[\int_0^{\infty} \frac{\lambda_2}{2} \; e^{-\lambda_2 y'} F\left(\frac{y'}{E_2}\right) dy'\right] \qquad (6\text{-}18)$$

Secondly, V_1 may be finally selected if V_1 and V_3 are selected on X_1 and then V_1 is selected when it is superior to V_2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\text{Prob.}(V_1 \text{ and } V_3 \text{ are selected on } X_1) = P(\bar{x}_{12}-\bar{x}_{13} < 0, \bar{x}_{12}-\bar{x}_{11} < 0)$$

$$= F_{2,\frac{1}{2}}\left[-\frac{(u_{12}-u_{13})}{E}, -\frac{(u_{12}-u_{11})}{E}\right]$$

$$= F_{2,\frac{1}{2}}\left[\frac{-y_2}{E}, \frac{-y_2+y_1}{E}\right] \qquad (6\text{-}19)$$

Where $y_2 = u_{12}-u_{13}$ and $y_1 = u_{11}-u_{13}$ and they are greater than zero. Density function for y_1 and y_2 is given in (6-16) Thus integrating (6-19) with respect to (6-16) will give the joint probability that V_3 is inferior to V_2 (no relation between V_1 and V_2 is assumed) and V_1 and V_3 are selected. It is,

$$2\int_0^{\infty}\int_0^{y_2} \frac{\lambda_1^2 e^{-\lambda_1 y_1-\lambda_1 y_2}}{3} F_{2,\frac{1}{2}}\left[\frac{-y_2}{E}, \frac{-y_2+y_1}{E}\right] dy_1\, dy_2 \qquad (6\text{-}20)$$

Further, the probability of V_1 being selected, given that V_1 and V_3 have been selected on X_1 (assuming no relation between u_{21} and u_{23}) is independent of n and is equal to $\frac{1}{2}$.

Probability that V_1 is greater than V_2 will be given by integrating (6-16) over the positive range and it comes to be $\frac{1}{2}$. From the above stipulations the joint probability of V_1 and V_2 being superior to V_3 in X_1, V_1 is superior to V_2 in X_2, V_1 and V_3 are selected on X_1 and finally V_1 is selected on X_2 is given by multiplying (6-20) by 1/4.

$$\frac{2}{4}\int_0^{\infty}\int_0^{y_2} \frac{\lambda_1^2 e^{-\lambda y_1}}{3} e^{-\lambda_1 y_2} F_{2,\frac{1}{2}}\left[\frac{-y_2}{E}, \frac{-y_2+y_1}{E}\right] dy_1 dy_2 \qquad (6\text{-}21)$$

Probability of correctly selecting V_1 under the above assumption will be given by summing (6-18) and (6-21) and a little algebra shows it to be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{\left[\frac{1}{2}+e^{\frac{\lambda_2^2E_2^2}{2}}I(\lambda_2E_2)\right]}{2}\times\frac{2}{3}\Big|\tfrac{1}{2}F_{2,\frac{1}{2}}(0,0)+\int_0^\infty\int_{-\infty}^0 e^{-\lambda_1E_1y_2}f_{2,\frac{1}{2}}(y_1y_2)dy_1dy_2$$

$$-\int_0^\infty\int_{-\infty}^{y_2} e^{-\lambda_1E_1y_2}f_{2,\frac{1}{2}}(y_1y_2)dy_1dy_2+\int_0^\infty\int_0^{y_2} e^{-\lambda_1E_1y_2}\,e^{-\lambda_1E_1y_1}f_{2,\frac{1}{2}}(y_1y_2)dy_1dy_2$$

$$+\int_0^\infty\int_{-\infty}^{y_2} e^{-2\lambda_1E_1y_2}f_{2,\frac{1}{2}}(y_1y_2)dy_1dy_2)\Big]+\frac{2}{4\cdot3}\Big|\tfrac{1}{2}F_{2,\frac{1}{2}}(0,0)$$

$$-2\int_0^\infty\int_{-\infty}^{\infty-y_2} e^{-\lambda_1E_1y_2}f_{2,-\frac{1}{2}}(y_1y_2)dy_1dy_2+\tfrac{1}{2}\int_0^\infty\int_{-\infty}^{0} e^{-2\lambda_1E_1}f_{2,-\frac{1}{2}}(y_1y_2)dy_1dy_2$$

$$+\tfrac{1}{2}\int_0^\infty\int_{-y_2}^{0} e^{-2\lambda_1E_1y_2}f_{2,-\frac{1}{2}}(y_1y_2)dy_2-\tfrac{1}{2}\int_0^\infty\int_{-\infty}^{-y_2} e^{-\lambda_1E_1y_2}f_{2,-\frac{1}{2}}(y_1y_2)dy_1dy_2\Big] \tag{6-22}$$

The above expression goes to 1/6 as n goes to infinity and it goes to 1/18 as n approaches zero. The probability of correct selection will be given by summing over all permutations and there will be 6 in all. Because λ_1's are equal and also λ_2's are equal all the permutations will have equal values. Hence the probability of correct selection will be (6-22) multiplied by six and it is unity as n goes to infinity and 1/3 as n goes to zero, i.e., random selection.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PART II

SCREENING PROBLEM

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 7

FIRST STAGE SCREENING

I. Screening problem

In the first part we have discussed the selection problem. The principal theme of the selection problem is that we are given a number (k) of varieties out of which the best or 't' best are to be selected for final adoption. This type of problem is posed, generally, at the final stage of the breeding programme and one of its main features is that k will not be very large; of the order of 10 to 15.

The whole breeding programme begins with the production of new varieties. Normally, many hundred, if not thousands, of new varieties will be produced in a single cohort and experience shows that a large proportion of these will not come up to the required standards. So the problem at this stage is to skim off relatively better varieties and this process is known as screening. Whereas, in selection the interest is in the best variety, in screening one is interested in all the varieties which are better than some set standard. Screening is one of the principal means of improving the quality of the varieties. We say that the quality of the varieties has improved if the average yielding capacity of the selected varieties is better than the average yielding capacity of the whole cohort: The same idea can be conveyed by saying that the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

population of varieties (selected fraction) has gained in the average yielding capacity by virtue of screening. The increase in the average yielding capacity may be called the 'gain'.

To form a basis of selection (or rejection) it will be desirable to introduce the standard, i.e. a variety which has already been established and select only those which exceed the standard by a certain specified amount. By replicating the standard more than the others (twice as much as others or even more) the precision of the experiment can be increased considerably. Increased replications of the standard will be particularly useful when the number of varieties is so large that the replications will be limited in number.

The whole cohort may be assumed as consisting of two groups.

1. One group is inferior to the standard variety

2. The other group is superior to the standard variety.

In practice it will be desirable to designate a variety superior to the standard only if it exceeds the standard in its yielding capacity by at least a certain specified amount δ, of course δ will have to be defined by the experimenter taking into consideration all the factors relevant to the particular selection problem. Ideal procedure will be that by means of which all the superior - as defined above - varieties are selected at every stage and the inferior varieties are rejected at the earliest possible. But that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is obviously impossible to attain. In any procedure adopted there are bound to be errors of both kind; some of the superior varieties will be wrongly rejected and some of the inferior ones will be wrongly accepted for the next stage.

If the yielding capacity of each variety is known there is no problem; the two groups can be separated easily. However, the yielding capacities are not known but we can perform an experiment to know their yielding capacities. The yield as depicted by the experiment will not be the true yielding capacity but it will contain an element of error. If we increase the replications to a very large number the selection will approach the ideal one but that increases the cost of the experiment. So a balance has to be struck between the two objectives. Finally, it is reasonable to assume that the experimenter will settle for a given probability, $1 - \alpha$, with which the superior variety will be selected.

In a two stage selection the probability, $1 - \alpha$, can be attained in an infinite combinations of $(1-\alpha')$ and $(1-\alpha'')$ such that $(1-\alpha')(1-\alpha'') = 1-\alpha$; $1-\alpha'$ is the probability that the superior variety will be selected at the first stage and $1-\alpha''$ is the probability that the superior variety will be selected at the second stage. At the second stage two alternatives are possible; (a) the yield of the second stage only is considered and (b) the combined yields of the first and the second stage are considered.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The procedure adopted for selection will be based on sample means of the varieties. Those varieties which exceed the standard by a fraction of δ, say, $(a \cdot \delta)$, are selected; At the second stage 'a' may be increased. For given error variance the number of replications required to have the desired probability can be determined as follows. Firstly, the notation used in sequel is introduced.

V_i superior implies mean of $V_i >$ mean $V_o + \delta$, V_o is the standard variety

$\bar{X}_i$ = average yield at first stage of variety V_i

$\bar{\bar{X}}_i$ = average yield at second stage of V_i

n_1 = sample size (replications) for the first stage

n_2 = sample size for the second stage

$n = n_1 + n_2$

I.1.

Error variance σ^2 is known. Probability that the superior variety is selected will be minimum when it is just δ superior to the standard i.e. mean yielding capacity of the variety V_i is just δ greater than the standard. So we calculate for n_1 such that the probability of selection of the superior variety is at least $1-\alpha$.

(a) First Stage.

$$\Pr.(V_i \text{ sup is selected}) = P(\bar{X}_i - \bar{X}_o > a\delta)$$

$$= P(Z_{1i} > h) = 1-\alpha' \quad (7\text{-}1)$$

$$h = \frac{(a-1)\delta}{\sigma} \sqrt{\frac{n_1}{2}}$$

Z_{1i} is normal with zero mean and unit variance. From

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(7-1) n_1 can be found in a straightforward manner.

(b) Second Stage.

If the result of the second stage alone is considered the procedure will be just the same. However, if the combined result of the first and the second stage is considered we may proceed as follows:

$$P(V_i \text{ sup selected at second stage} \mid \text{selected at 1st stage})$$
$$= P(\bar{\bar{X}}_i - \bar{\bar{X}}_o > b\delta \mid \bar{X}_i - \bar{X}_o > a\delta) = P(Z_{2i} > k \mid Z_{1i} > h) = 1-\alpha'' \tag{7-2}$$

h is already defined and $k = \frac{(b-1)\delta}{\sigma}\sqrt{\frac{n}{2}}$

Z_{2i} and Z_{1i} are both N(0 1) with correlation $\sqrt{n_1/n}$.

This will determine a bivariate normal distribution, tables for that are available. However, the tables are given for certain discrete values of correlation r, and for r, h and k differing from the tabular values interpolation is required. If interpolation has to be carried out in more than one variable it can be pretty tedious. We can avoid bivariate integral and can get pretty good approximation.

$$Z_{1i} = \frac{(\bar{X}_i - \bar{X}_o - \delta)}{\sigma}\sqrt{\frac{n_1}{2}} \quad \text{and it is N(0,1)}$$

$$E(Z_{1i} \mid Z_{1i} > h) = \frac{1}{\sqrt{(2\pi)}} \int_h^\infty X\, e^{-\frac{X^2}{2}}\, dX \Big/ _{1-\alpha} \quad \text{by virtue of (7-1)}$$

$$= c\text{-say},$$

That will give

$$E\left[\sqrt{\frac{n_1}{2}}\,\frac{(\bar{X}_i - \bar{X}_o - \delta)}{\sigma} \Big/ Z_{1i} > h\right] = c$$

$$\text{or} \quad E(\bar{X}_i - \bar{X}_o \mid Z_{1i} > h) = \sigma \cdot c \Big/ \sqrt{\frac{n_1}{2}} + \delta \tag{7-3}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\bar{\bar{X}}_i - \bar{\bar{X}}_o = \frac{1}{n}\left[n_1(\bar{X}_i - \bar{X}_o) + (X_i^{n_1+1} + \ldots + X_i^{n}) - (X_o^{n+1} + \ldots + X_o^{n})\right]$ combined with (7-3) it will give

$$E(\bar{\bar{X}}_i - \bar{\bar{X}}_o \mid Z_{1i} > h) = \frac{n_1 . c\sigma}{n\sqrt{n_1/2}} + \delta \qquad (7\text{-}4)$$

Thus $P(V_i$ sup selected at II stage / selected at I stage)

$$= P(\bar{\bar{X}}_i - \bar{\bar{X}}_o > b\delta \;/\; \bar{X}_i - \bar{X}_o > a\delta)$$

$$= \frac{P(Z_{2i} > k')}{P(\bar{X}_i - \bar{X}_o > a\delta)} = 1-\alpha'' \qquad (7\text{-}5)$$

where $k' = \left[\delta(b-1) - \frac{\sqrt{2}c\sigma n_1}{\sqrt{n_1 n}}\right]\sqrt{\frac{n}{2\sigma}}$

Z_{2i} has zero mean and unit variance but it is not normal. However by taking it as a normal variate we can get a fairly good approximation for $1-\alpha''$ and hence for $1-\alpha$. $\bar{X}_i - \bar{X}_o$ is a component of Z_{2i}. From stage I we know that $\bar{X}_i - \bar{X}_o > a\delta$ but by taking Z_{2i} as normal with zero mean and unit variance we do not give due weight to this component of Z_{2i}. Therefore, (7-5) will give the lower limit for $1-\alpha''$. By taking $E(\bar{\bar{X}}_i - \bar{\bar{X}}_o) = \frac{c\sigma}{\sqrt{n_1/2}} + \delta$ instead of $(\frac{n_1}{n})\frac{c.\sigma}{\sqrt{n_1/2}} + \delta$ we give more weight to the $\bar{X}_i - \bar{X}_o$ component of Z_{2i} and that will give upper limit of (7-5).

I.2.

Error variance σ^2 is unknown but an estimate s^2 of variance is known or can be obtained.

(a) First Stage.

$$P(V_i \text{ sup selected}) = P(\bar{X}_i - \bar{X}_o > a\delta)$$

$$= P\left(\frac{Z_{1i}}{s} > h\right) = 1-\alpha' \qquad (7\text{-}6)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Z_{1i} is $N(0,\sigma^2)$ and $h = \frac{\delta(a-1)}{s}\sqrt{\frac{n_1}{2}}$

Thus Z_{1i}/s follows t distribution. From t tables we can easily find h and hence n_1 for given $1-\alpha'$.

(b) Second Stage.

Again if the result of the second stage alone is considered there is no problem. In case the result of first stage is also considered we have,

$$P(V_i \text{ sup selected}) = P(\bar{\bar{X}}_i - \bar{\bar{X}}_o > b\delta \mid \bar{X}_i - \bar{X}_o > a\delta)$$

$$= P\left\{Z_2/s > \delta\frac{(b-1)}{s} \mid Z_1/s > \frac{(a-1)}{s}\delta\right\} \tag{7-7}$$

Z_2 is $N(0,\frac{2}{n}\sigma^2)$ and Z_1 is also $N(0,\frac{2}{n_1}\sigma^2)$ correlation of Z_2 and Z_1 is $\sqrt{\frac{n_1}{n}}$ Joint density of Z_2, Z_1 and s is

$$\frac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi\, 2^{\frac{N}{2}-1}\Gamma(\frac{N}{2})\gamma\beta}\, N^{N/2}\, \frac{s^{N-1}\text{Exp}\left|-\frac{1}{2(1-\rho^2)}\left(\frac{Z_1^2}{\beta^2\sigma^2} + \frac{Z_2^2}{\gamma^2\sigma^2} - \frac{2\rho Z_1 Z_2}{\beta\,\gamma\,\sigma^2}\right)\right|}{\sigma^{N+2}}\, e^{\frac{-Ns^2}{2\sigma^2}} \tag{7-8}$$

N is the number of degrees of freedom for s

Substitute $Z_1 = t_1 s$, $\beta = \sqrt{2/n_1}$ $\gamma = \sqrt{2/n}$

$Z_2 = t_2 s$ and integrating out s will give the joint density function of t_1 and t_2 and it is

$$\frac{\Gamma(\frac{N+2}{2})}{N\pi(1-\rho^2)^{\frac{1}{2}}\Gamma(\frac{N}{2})\gamma\beta}\left[1 + \frac{\frac{t_1^2}{\beta^2} + \frac{t_2^2}{\gamma^2} - \frac{2\rho t_1 t_2}{\beta\gamma}}{(1-\rho^2)(N)}\right]^{-(N+2)/2} \tag{7-9}$$

To get the value of $1-\alpha$ we need to integrate for $t_1 > \delta\frac{(a-1)}{s}$ and for $t_2 > \delta\frac{(b-1)}{s}$

(7-9) is very similar to the bivariate generalization of t distribution as worked out by Dunnett (16). The table given there is for correlation equal to 1/2 and for equidistant

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

points only. Moreover the two variables have equal variances. The integration of (7-9) is pretty tedious. Following the procedure of section II for the case when the variance is known we can have fairly good approximation for this case too and (7-7) reduces to

$$\frac{P(\frac{Z_2}{s} > k')}{P(\frac{Z_1}{s} > h)} = 1-\alpha'' \quad k' = \left[(b-1)\delta - \frac{c.s}{\sqrt{\frac{n_1}{2}}} \frac{n_1}{n} \right] \frac{1}{s} \sqrt{\frac{n}{2}}$$

$$h = \frac{(a-1)}{s}\delta \sqrt{\frac{n_1}{2}} \qquad (7\text{-}10)$$

In (7-10) numerator as well as the denominator can be found from t tables. Moreover at the II stage the variance can be assumed known and then the results of previous section can be used.

II.

As stated earlier screening is one of the several means of improving the quality of the population (of varieties). The gain in quality - increase in the yielding capacity - that can be achieved by screening alone should be compared with possible gain that could be obtained by any other device or it may be better to adopt screening in conjunction with other means. The first step towards that will be to know the gain possible by adopting screening alone. In these sections we evaluate the one stage consequences of the procedure outlined in the previous section as affecting the gain. The gain is obtained in units of standard deviation of the distribution we work with and as such is termed standard gain.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Obviously, the gain will depend upon the distribution function of the varietal means. Generally, the parent population is assumed to be normal - it is easy to work with. How far the assumption of normality is justified in a particular situation can be decided empirically only. Even if the parent population is assumed to be normal, the operations performed preceding the selection would have resulted in introducing skewness to the distribution. Consequently, it will be desirable to see how does the departure from normality affect the consequences of the procedure adopted. In the following sections we evaluate the consequences of single stage selection for normal, exponential and double exponential distributions.

II.1. Normal.

The distribution of varietal means is normal with mean u and variance σ_o^2 and the error variance is σ^2. According to the procedure described earlier --

$$P(V_i \text{ selected}) = P(\bar{X}_i - \bar{X}_o > a\delta)$$

That will give $P(V_i \text{ with mean } Y \text{ selected}) = \frac{1}{\sqrt{2\pi}\, E} \int_{a\delta}^{\infty} \exp -\tfrac{1}{2}\left(\frac{X-Y+u_o}{E^2}\right)^2 dX$

$$= I\frac{(\bar{a}-Y)}{E}$$

Where $E = \sqrt{(2/n)}\sigma$; u_o is the mean of the standard variety and $\bar{a} = a\delta + u_o$. Y is supposed to be normal with mean u and variance σ_o^2. That will give the fraction of the population selected as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{1}{2\pi\, E\, \sigma_o} \int_{-\infty}^{\infty}\int_{a\delta}^{\infty} \exp -\tfrac{1}{2}\left[\frac{(Y-u)^2}{\sigma_o^2} + \frac{(X-Y+u_o)^2}{E^2}\right] dX\,dY \qquad (7\text{-}11)$$

Substitute $Z_1=\frac{(X-u+u_o)}{\sigma}\sqrt{\frac{n_1}{2}}$ and $Z_2=\frac{Y-u}{\sigma_o}\sqrt{\frac{n_1}{2}}$ and (7-11) reduces to

$$\sqrt{\frac{2}{n_1}}\,\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{\frac{(\bar{a}-u)}{\sigma}\sqrt{\frac{n_1}{2}}}^{\infty} \exp -\tfrac{1}{2}\left| Z_2^2\left(\frac{1}{n_1}+\frac{\sigma_o^2}{\sigma^2}\right)+Z_1^2-2Z_1Z_2\,\frac{\sigma_o}{\sigma_1}\right| dZ_1\,dZ_2 \qquad (7\text{-}12)$$

(7-12) is a bivariate integral with variance covariance matrix

Z_1	Z_2
$1+\frac{n_1 m^2}{2}$	$\frac{n_1 m}{2}$
$\frac{n_1 m}{2}$	$\frac{n_1}{2}$

$m=\frac{\sigma_o}{\sigma}$

Thus (7-12) can be written as

$$\frac{1}{\sigma_{Z_1}\sigma_{Z_2}\,2\pi(1-\rho^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{(\frac{a-u}{\sigma})\sqrt{\frac{n_1}{2}}}^{\infty}\exp -\tfrac{1}{2}\left(\frac{Z_1^2}{\sigma_{Z_1}^2}+\frac{Z_2^2}{\sigma_{Z_2}^2}-\frac{\rho 2 Z_1 Z_2}{\sigma_{Z_1}\sigma_{Z_2}}\right)\frac{1}{(1-\rho^2)}\, dZ_1\,dZ_2 \qquad (7\text{-}13)$$

Changing to standardized variables and denoting the standardized variables by the same letters (7-13) reduces to

$$\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{t}^{\infty}\exp -\tfrac{1}{2}(Z_1^2+Z_2^2-2\rho Z_1 Z_2)/(1-\rho^2)\; dZ_1\,dZ_2 \qquad (7\text{-}14)$$

Integrating out Z_2 we get

$$\frac{1}{\sqrt{2\pi}}\int_{t}^{\infty} e^{-Z_1^2/2}\, dZ_1 = P, \text{ say}$$

$$t = \frac{(\bar{a}-u)}{\sqrt{E^2+\sigma_o^2}}\,\frac{1}{}$$

Z_2 has a regression on Z_1 and as Z_1 and Z_2 are standardized variables the regression will equal the correlation coefficient determined from the covariance matrix given above.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The mean of Z_1 after truncation at 't' is $\lambda(t)$ and $\lambda(t)=\frac{1}{\sqrt{2\pi}}\ e^{-t^2/2}\Big/ I(t)$. By virtue of the regression of Z_2 on Z_1 the mean of Z_2 after truncation of Z_1 at t will be

$$\frac{\sigma_o}{\sqrt{E^2+\sigma_o^2}}\qquad \frac{\frac{1}{\sqrt{2\pi}}\ e^{-t^2/2}}{I(t)} \tag{7-15}$$

$Z_2=\frac{Y-u}{\sigma_o}$, hence (7-15) gives the gain in the yielding capacity (improvement in quality) in terms of the standard deviation of the parent population.

Exponential

II.2.

In this section the parent population is assumed to be exponential with parameter λ and the results corresponding to normal distribution are obtained for the first stage selection.

Fraction selected will be

$$\int_o^\infty \lambda\ e^{-\lambda Y}\ I(\frac{\bar{a}-Y}{E})dY = \int_o^\infty \lambda e^{-\lambda Y} F(\frac{Y-\bar{a}}{E})dY$$

$$= F(\frac{-\bar{a}}{E}) + \frac{1}{E}\int_o^\infty e^{-\lambda Y} f(\frac{Y-\bar{a}}{E})dY$$

$$= F(\frac{-\bar{a}}{E}) + e^{\frac{1}{2}(\lambda^2E^2-2\lambda\bar{a})} I(\frac{\lambda E^2-\bar{a}}{E}) \tag{7-16}$$

To have comparable results with the normal we work in terms of standard units or it may be defined as standard gain. For that the mean of $\frac{Y-1/\lambda}{1/\lambda}$ after selection is required. Hence the standard gain for exponential is

$$\int_o^\infty (\lambda Y-1)\lambda\ e^{-\lambda Y}\ F(\frac{Y-\bar{a}}{E})dY \Big/ P$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$= \int_0^\infty \lambda^2 \bar{e}^{\lambda Y} Y . F\left(\frac{Y-\bar{a}}{E}\right) \Big/ P-1 = \int_0^\infty \lambda \bar{e}^{\lambda Y} Y \; f\left(\frac{Y-\bar{a}}{E}\right) dY \Big/ P$$

$$= e^{\frac{1}{2}(\lambda^2 E^2 - 2\lambda\bar{a})} \left[\frac{\lambda E \cdot e^{-\frac{(\lambda E^2-\bar{a})^2}{2E^2}}}{\sqrt{2\pi}} + \lambda(\bar{a}-\lambda E^2) I\left(\frac{\lambda E^2-\bar{a}}{E}\right) \right] \Big/ P \qquad (7\text{-}17)$$

II.3. Double Exponential.

In this section the population is assumed to be double exponential with parameters λ and u. Double exponential is characterized by the density function

$$\tfrac{1}{2} \bar{e}^{\lambda |Y-u|} \qquad -\infty < Y < \infty$$

Prob(that V_i with mean Y is selected) $= F\left(\frac{Y-\bar{a}}{E}\right)$

That will give for the double exponential the fraction of the population selected as

$$\tfrac{1}{2} \int_{-\infty}^{\infty} \lambda \bar{e}^{\lambda |Y-u|} F\left(\frac{Y-\bar{a}}{E}\right) dY$$

$$= \tfrac{1}{2} \left| \lambda \int_u^\infty \bar{e}^{\lambda(Y-u)} F\left(\frac{Y-\bar{a}}{E}\right) dY + \lambda \int_{-\infty}^u \bar{e}^{\lambda(Y-u)} F\left(\frac{Y-\bar{a}}{E}\right) dY \right| = P \qquad (7\text{-}18)$$

(7-18) may be written as $\tfrac{1}{2} \left| (A)+(B) \right|$

$$(B) = \lambda \int_{-\infty}^u \bar{e}^{\lambda(Y-u)} F\left(\frac{Y-\bar{a}}{E}\right) dY = I\left(\frac{\bar{a}-u}{E}\right) - e^{\frac{\lambda^2 E^2}{2} + \lambda(\bar{a}-u)} F\left(\frac{u-\bar{a}-\lambda E}{E}\right)$$

$$(A) = \int_u^\infty \lambda \; e^{\lambda u - \lambda Y} F\left(\frac{Y-\bar{a}}{E}\right) dY = I\left(\frac{\bar{a}-u}{E}\right) + e^{\frac{\lambda^2 E^2 + \lambda(u-\bar{a})}{2}} I\left(\frac{u-\bar{a}+\lambda E}{E}\right)$$

Combining the two results we get the fraction selected as

$$\tfrac{1}{2} \left[2I\left(\frac{\bar{a}-u}{E}\right) + e^{\frac{\lambda^2 E^2}{2} + \lambda(u-\bar{a})} I\left\{ \left(\frac{u-\bar{a}}{E}\right) + \lambda E \right\} - e^{\frac{\lambda^2 E^2 + (\bar{a}-u)}{2}} F\left(\frac{u-\bar{a}-\lambda E}{E}\right) \right] \qquad (7\text{-}19)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Variance of double exponential is $2/\lambda^2$ that will give the standard gain as $\frac{E(Y-u)}{\sqrt{2}/\lambda} = \lambda E(Y-u)/\sqrt{2}$.

$$= \frac{\lambda}{2\sqrt{2}} \int_{-\infty}^{\infty} (Y-u)\lambda e^{-\lambda|Y-u|} F\left(\frac{Y-\bar{a}}{E}\right) dY \Big/ P$$

$$= \frac{1}{P.2\sqrt{2}}\left[\lambda^2 \int_{u}^{\infty} e^{\lambda u-\lambda Y} F\left(\frac{Y-\bar{a}}{E}\right) dY + \lambda^2 \int_{-\infty}^{u} Y e^{-\lambda u+\lambda Y} F\left(\frac{Y-\bar{a}}{E}\right) dY\right] - \frac{\lambda u}{\sqrt{2}} \qquad (7\text{-}20)$$

(7-20) may be written as $\frac{1}{P.2\sqrt{2}}\left[(C)+(D)\right] - \frac{\lambda u}{\sqrt{2}}$

Solving (C) and (D) separately and substituting back we will get the standard gain for the double exponential

$$\frac{1}{2\sqrt{2}P}\left[2\lambda u\, I\left(\frac{\bar{a}-u}{E}\right) + e^{\frac{\lambda^2E^2+\lambda(u-\bar{a})}{2}} I\left\{\left(\frac{u-\bar{a}}{E}\right)+\lambda E\right\} + F\left(\frac{u-\bar{a}}{E}-\lambda E\right) e^{\frac{\lambda^2E^2+\lambda(\bar{a}-u)}{2}}\right]$$

$$+ \frac{1}{2\sqrt{2}P.}\left[e^{\frac{\lambda^2E^2-\lambda(\bar{a}-u)}{2}}\left\{\frac{\lambda E}{\sqrt{2\pi}} e^{\frac{(u-\bar{a}+\lambda E^2)^2}{2E^2}} + \lambda(a-\lambda E^2) I\left(\frac{u-\bar{a}+\lambda E^2}{E}\right)\right]\right.$$

$$\left. + e^{\frac{\lambda^2E^2+\lambda(\bar{a}-u)}{2}}\left\{\frac{\lambda E}{\sqrt{2\pi}} e^{-\lambda(u-\bar{a}-\lambda E^2)^2/2E^2} - \lambda(\bar{a}+\lambda E^2) F\left(\frac{u-\bar{a}}{E}-\lambda E\right)\right\}\right] - \frac{\lambda u}{\sqrt{2}} \qquad (7\text{-}21)$$

Numerical Results

III.

Having obtained the formulae for the fraction selected and the gain obtained it will be desirable to have numerical results and see how they differ in the three distributions. Tables 10, 12 and 14 give the fraction selected for the normal, exponential and the double exponential respectively and tables 11, 13 and 15 give the standard gain for the normal exponential and double exponential respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Firstly it will be in order to explain how the tables have been constructed. As explained earlier a variety is selected if it exceeds the standard by a certain amount 'aδ' i.e.

$$\bar{X}_i - \bar{X}_o > a\delta$$

$$\text{or } \bar{X}_i - \bar{X}_o - (u_i - u_o) > \bar{a} - u_i$$

The probability of a variety to be selected depends upon how far '$\bar{a}$' and u_i are situated and their relative position i.e. whether '$\bar{a}$' is on the left of u_i or on its right. Likewise the consequences of the procedure with respect to a given distribution will depend upon the raltive position of '$\bar{a}$' and u - the mean of the parent population. '$\bar{a}$' is taken greater than u and the difference of '$\bar{a}$' and u is taken in terms of E ($\sqrt{2}\sigma/\sqrt{n}$); where σ^2 is the error variance.

A number of values of $\frac{\sigma_o}{E}$ - (σ_o^2 is the variance of the parent population) - have been taken and corresponding to each ratio, the fraction selected and the gain obtained have been evaluated for different values of $\frac{\bar{a}-u}{E}$. If it is required ($\frac{\bar{a}-u}{E}$) can be easily changed into $\frac{\bar{a}-u}{\sigma_o}$; ($\frac{\bar{a}-u}{\sigma_o} = \frac{\bar{a}-u}{E}\frac{E}{\sigma_o}$).

This change simply means that the scale on the vertical side of the table is changed by a constant. Therefore for comparison between different distributions any one can be used; $\frac{\bar{a}-u}{E}$ has been adopted in these tables for the sake of convenience in the formulae employed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Normal Distribution.

Table 10

Fraction Selected

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.1	.2	.3	.4	.5	.6	.8	1.0	1.5
.1	.4603	.4609	.4619	.4630	.4643	.4659	.4688	.4718	.4778
.2	.4210	.4222	.4240	.4263	.4290	.4319	.4379	.4438	.4558
.3	.3827	.3842	.3869	.3903	.3942	.3985	.4074	.4160	.4339
.5	.3092	.3118	.3160	.3212	.3273	.3340	.3481	.3745	.3981
.75	.2275	.2310	.2362	.2431	.2511	.2605	.2790	.2979	.3387
1.0	.1596	.1632	.1691	.1766	.1855	.1955	.2173	.2397	.2896
1.50	.06763	.0706	.0755	.0818	.0888	.0991	.1319	.1444	.2054
2.00		.02507	.0278	.---	.03692	---	---	.07865	.13395

Table 11

Standard Gain

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.1	.2	.3	.4	.5	.6	.8	1.0	1.5
.1	.0858	.1689	.2469	.3836	.4389	.5301	.5301	.5921	.7034
.2	.0924	.1818	.2653	.3409	.4092	.4683	.5786	.6249	.7245
.3	.0991	.1950	.2842	.3644	.4365	.4982	.5943	.6582	.7552
.5	.1131	.2225	.3233	.4133	.4931	.5660	.6637	.7271	.8031
.75	.1319	.2585	.3748	.4786	.5671	.6406	.7527	.8225	.8996
1.00	.1514	.2962	.4284	.5441	.6445	.7266	.8454	.9163	.9839
1.50	.1923	.3753	.5412	.6397	.8162	.9049	1.0122	1.1130	1.1535

In the case of normal running down the columns of table 10 it is clear that as $\frac{\bar{a}-u}{E}$ is increased the fraction selected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

decreases and that is what can be expected; by increasing the cut-off point selection becomes more intense and consequently a smaller fraction is selected. Viewing horizontally, as the value of $\frac{\sigma_o}{E}$ increases fraction selected increases. This can be viewed as this; if we take the value of E fixed the increasing of $\frac{\sigma_o}{E}$ implies an increase in σ_o, i.e., parent distribution is more flat and that explains the greater percentage selected.

Table 11 depicts the standard gain for the normal. It shows an increase in gain downward as well as horizontally to the right. Downward increase in gain is obvious, as the selection becomes more intense gain increases. The increased gain to the right can be explained by the increased flatness of the parent distribution with increasing values of $\frac{\sigma_o}{E}$.

Exponential

Table 12

Fraction selected

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.2	.3	.5	1.0	1.5
.1	.4612	.4589	.4540	.4319	.4096
.2	.4216	.4212	.4181	.4027	.3860
.3	.3856	.3848	.3835	.3744	.3638
.5	.3114	.3204	.3173	.3212	.3219
.75	.2307	.2345	.2431	.2616	.2750
1.0	.1631	.1683	.1804	.2104	.2342
1.5	.0710	.0761	.0904	.1325	.16735
2.0	.0253	.02902	.04065	.0822	.1211

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 13

Standard gain for the exponential

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.2	.3	.5	1.0	1.5
.1	.1593	.2385	.3594	.5730	.7200
.2	.1823	.2744	.4143	.6249	.7722
.3	.1980	.2906	.4365	.6801	.8259
.5	.2360	.3583	.5249	.7989	.9399
.75	.2885	.4295	.6541	.9646	1.0859
1.0	.3469	.6066	.8065	1.1483	1.2399
1.5	.4891	.7787	1.1940	1.5617	1.5931

The general behaviour in the exponential is somewhat different than that in the normal. Running down the columns we find that the fraction selected decreases as the cut-off point is increased. But running horizontally to the right the fraction selected, unlike the normal, decreases at lower values of $\frac{\bar{a}-u}{E}$ and at higher values of $\frac{\bar{a}-u}{E}$ the fraction selected increases with an increase in $\frac{\sigma_o}{E}$. The turning point is near $\frac{\bar{a}-u}{E}$ equal to .5.

The relative difference in the fraction selected between the normal and the exponential is as follows. At lower values of $\frac{\sigma_o}{E}$ the fraction selected is almost the same for the two distributions. The values calculated agree up to two decimal places. Normal distribution gives a little greater fraction, than the exponential, though the difference is at the third decimal place. This is at lower values of $\frac{\bar{a}-u}{E}$; at higher values of $\frac{\bar{a}-u}{E}$ the positions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

are reversed i.e. the exponential gives a higher fraction, though the difference is again at the third decimal place.

At higher values of $\frac{\sigma_o}{E}$ ($\frac{\sigma_o}{E} = 1.5$) the normal gives much higher values of the fraction selected than the exponential. The difference is up to 4%. However as the value of $\frac{\bar{a}-u}{E}$ is increased the difference between the two decreases and eventually exponential exceeds, though the difference is not much.

The general inference from this may be drawn that if $\frac{\sigma_o}{E}$ is not very large (< 1.0) and the cut-off point is such that about 1/10 or less of the population is selected there would not be much difference in the fraction selected for the one stage selection for the two cases.

Regarding the standard gain obtained on account of the selection procedure, except at low values of $\frac{\sigma_o}{E}$ and $\frac{\bar{a}-u}{E}$ where normal exceeds the exponential by a small margin, exponential gives greater gain than the normal and in the range studied the difference increases with the increase of $\frac{\bar{a}-u}{E}$ and $\frac{\sigma_o}{E}$.

Double Exponential.

Table 14

Fraction selected

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.2	.3	.4	.5	.6	.8	1.0	1.5
.1	.4719	.4667	.4675	.4661	.4703	.4741	.4723	.4796
.2	.4336	.4240	.4261	.4285	.4319	.4353	.4396	.4494
.3	.3844	.3866	.3904	.3932	.3959	.4037	.4102	.4276
.5	.3128	.3190	.3200	.3257	.3318	.3417	.3525	.3817
.75	.2319	.2467	.2421	.2491	.2562	.2636	.2860	.3195
1.0	.163	.1658	.1759	.1837	.1898	.2191	.2217	.2519
1.5	--	.0743	.0767	.0887	.0969	.1148	.1336	.1785

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 15

Standard Gain for the double exponential

$\frac{\bar{a}-u}{E}$ \ $\frac{u}{\sigma_o}$	$\sigma_o/E = .5$			$\sigma_o/E = 1.0$		
	.5	1.0	1.5	.5	1.0	1.5
.1	.2826	.2945	.2913	.5207	.5181	.5151
.2	.3151	.3225	.3226	.5615	.5587	.5620
.3	.3385	.3348	.3348	.6092	.5911	.5922
.5	.4387	.3979	.3998	.6567	.6657	.6658
.75	.4533	.4439	.4336	.7577	.7489	.7402
1.0	.5276	.5437	.5265	.9583	.9867	.9503
1.5	.7916	.8224	.8790	1.1515	1.5128	1.1513

Apart from a few enteries at low values of $\frac{\bar{a}-u}{E}$ and $\frac{\sigma_o}{E}$ the general behaviour of the double exponential is like that of the normal distribution. The fraction selected increases horizontally to the right and decreases down the columns.

At low values of $\frac{\sigma_o}{E}$ (.2 & .3) the fraction selected in case of double exponential is a little greater than the normal. The difference is in the third decimal place. For $\frac{\sigma_o}{E}$ equal to .4 the fraction selected in the case of double exponential is little larger for low values of $\frac{\bar{a}-u}{E}$ and a little less at high values of $\frac{\bar{a}-u}{E}$. Up to $\frac{\sigma_o}{E} = .6$ the difference either way is very small. For $\frac{\sigma_o}{E}$ equal to .8 and 1.0 at higher values of $\frac{\bar{a}-u}{E}$ double exponential gives up to 1.1% less values than the normal. For $\frac{\sigma_o}{E} = 1.5$ the difference increases up to 4% and the double exponential

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is less.

The general inference is again the same that if $\frac{\sigma_o}{E}$ is not very large (<1.0) the fraction selected in normal as well as in double exponential will not differ appreciably. Combining the inference with that dereived in the case of the exponential it may be stated that for small values of $\frac{\sigma_o}{E}$ the normal, the double exponential and the exponential give similar values for the fraction selected.

In normal and exponential the standard gain like the fraction selected is a function of $\frac{\bar{a}-u}{E}$ and $\frac{\sigma_o}{E}$ only but, in the double exponential it depends upon $\frac{u}{\sigma_o}$ (ratio of the mean and the standard deviation of the distribution). Standard gain for two values of $\frac{\sigma_o}{E}$ (.5 & 1.0) has been calculated. For each value of $\frac{\sigma_o}{E}$ three ratios of u and σ_o (.5, 1.0 & 1.5) have been taken.

The gain is different with different values of $\frac{u}{\sigma_o}$ but, there does not appear to be any trend, i.e. whether higher values of $\frac{u}{\sigma_o}$ give less or more gain. The three ratios give nearly similar values, in some cases one exceeds the other and in others the position is reversed. The fluctuation is greater at lower values of $\frac{\sigma_o}{E}$. At lower values of $\frac{\bar{a}-u}{E}$ the gain is less for the double exponential and at higher values of $\frac{a-u}{E}$ the gain is larger for the double exponential. For $\frac{\sigma_o}{E} = 1.0$ double exponential exceeds the normal at $\frac{\bar{a}-u}{E} = 1.0$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

IV. Variance of gain

The gain has been worked out on the assumption that the means and variances of the distributions and the error variance are known, but, that is rarely the case. More likely the estimates of the parameters will be known or can be obtained by conducting an experiment. Below a procedure to work out the error of the gain from the experimental data is given. Design of the experiment may be assumed to be the simple randomized block design. A varietal trial with N varieties and n replications is conducted. Each recording of the yield can be assumed to be composed of true yielding capacity of the variety and an error element and the block effect, i.e.

$$X_{ij} = u_i + b_j + e_{ij}$$

$$= u + v_i + b_j + e_{ij}$$

X_{ij} is the observation on variety V_i on block j, u is the general mean, v_i the contribution of the variety V_i, b_j is the effect of the block j and e_{ij} is the error element.

From our basic assumption that the varietal means follow normal distribution with variance σ_o^2 it follows that v_i's are normal with the same variance and u can be specified in such a way that the sum of v_i's is zero; same holds for b_j; e_{ij}'s follow independently normal distribution with zero mean and known variance σ^2. With this terminology the analysis of variance can be written as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Source of variation	degrees of freedom	S.S.	M.S.S.	EMS
Mean	1			
Blocks	n-1	Q_3	$\frac{Q_3}{n-1}=g$	$\sigma^2+N\sigma_b^2$
Varieties	N-1	Q_2	$\frac{Q_2}{N-1}=p$	$\sigma^2+n\sigma_0^2$
Error	(n-1)(N-1)	Q_1	$\frac{Q_1}{(N-1)(n-1)}=f$	σ^2

From the above table we can get $\sigma_0^2 = (p-f)/n$. Further we have Var $(f) = 2(\sigma^2)^2/(n-1)(N-1)$ because Q_1 is $\sigma^2\chi^2$ with $(n-1)(N-1)$ degrees of freedom. Likewise on the basis that v_i's are distributed normally variance of $p=2(\sigma^2+n\sigma_0^2)^2/(N-1)$.

In the above analysis of variance table the control has not been indicated. The inclusion of the standard will not effect the analysis, however, its inclusion will lead to some petty changes in the constant factors, e.g., we will have N+1 varieties instead of N etc. To keep the things as straight and simple as possible the control has not been shown explicitly.

The estimate of varietal mean (over all mean) $y_{..} = \frac{\Sigma_{ij} X_{ij}}{nN}$

Each X_{ij} as noted before is $= u+v_i+b_j+e_{ij}$, from that we get the variance of the mean $y.. = \frac{\sigma_0^2+\sigma^2+\sigma_b^2}{nN}$, where σ_b^2 is the block variance. An estimate of $u_0 = \bar{X}_0$ with variance $\frac{\sigma_b^2+\sigma^2}{n}$

In the case of the exponential a priori distribution the estimate of the error variance σ^2 and block variance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

σ_b^2 will be obtained in the same way as in the case of normal, λ can be estimated by two independent estimates and those may be used to determine if the varietal means are really following an exponential distribution. The first estimate of λ is given by the mean (over all mean) of the varieties i.e.

$$y.. = \frac{\Sigma X_{ij}}{nN} = 1/\lambda$$

Second estimate of λ can be had from the analysis of variance table

$\sigma_o^2 = (p-f)/n$ and this is the same as

$1/\lambda^2 = (p-f)/n$

As in the normal case $\text{Var}(y..) = \dfrac{\sigma_b^2 + \frac{1}{\lambda^2} + \sigma^2}{nN}$

That is all the basic material we need to calculate the variance of the standard gain.

A function g of variables $x_1, x_2, \ldots x_k = g(x_1, x_2, \ldots, x_k)$ and $\dfrac{\partial g}{\partial m_i} = \dfrac{\partial g}{x_i}$ at $x_i = m_i$ where $m_i = E(x_i)$

then $\text{var}(g) = \sum_{i,j=1}^{k} \left| \dfrac{\partial g}{\partial m_i} \dfrac{\partial g}{\partial m_j} \text{Cov}(x_i x_j).. \right|$ ....(Kendall) (7-22)

i, and j may be equal in that case $\text{Cov}(x_i x_j) = \text{Var}(x_i)$

IV.1. Normal

Standard gain for the normal distribution as worked out in section II is

$$\sqrt{\frac{\sigma_o}{2\sigma^2 + \sigma_o^2}} \quad \frac{1}{\sqrt{2\pi}} \quad e^{-t^2/2} \Big/ \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-t^2/2} dt,$$ t has been

defined in section II. Substituting the equivalent quantities from the analysis of variance table the standard gain assumes

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the form

$$\frac{(p-f)^{\frac{1}{2}}}{P+f}\ \frac{1}{\sqrt{2\pi}}\ e^{-\frac{t^2}{2}} \Big/ \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{\frac{t^2}{2}}\,dt \quad \text{now } t=(a\delta+\bar{X}_o-Y..)\frac{\sqrt{n}}{\sqrt{P+f}}$$

We note that p, f, y.. and $\bar{X}_o$ are independent of each other. Thus using the formula in (7-13) we get the variance of the standard gain for the normal distribution.

$$\frac{(p-f)}{(p+f)^3}\ A^2\left(\frac{f}{p-f}+\frac{t^2-At}{2}\right)^2 \mathrm{Var}(p) + \frac{(p-f)}{(p+f)^3}\ A^2\left(\frac{t^2-At}{2}-\frac{p}{p-f}\right)^2 \mathrm{Var}(f)$$

$$+\frac{(p-f)n}{(p+f)^2}\ A^2(A-t)^2\ \mathrm{Var}(\bar{X}_o) + \frac{(p-f)^n}{(p+f)^n}\ A^2(t-A)^2\ \mathrm{Var}(y..) \qquad (7\text{-}13)$$

$$\text{where } A = \frac{1}{\sqrt{2\pi}}\ e^{-t^2/2} \Big/ \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-t^2/2}dt$$

Changing back to original variables (those may be assumed to be obtained from the experimental data rather than known a priori) and substituting the variances of p,f, etc. as worked out earlier the variance of the standard gain for the normal assumes the form.

$$\frac{n\sigma_o^2}{(n\sigma_o^2+2\sigma^2)^3}\ A^2\left[\frac{\sigma^2}{n\sigma_o^2}+\frac{t^2-At}{2}\right]^2 2\,\frac{(\sigma^2+n\sigma_o^2)^2}{N-1} \qquad \text{(i)}$$

$$+\frac{n\sigma_o^2}{(n\sigma_o^2+2\sigma^2)^3}\ A^2\left[\frac{t^2-At}{2}-\frac{n\sigma_o^2+\sigma^2}{n\sigma_o^2}\right]^2 \frac{2\sigma^4}{(n-1)(N-1)} \qquad \text{(ii)} \qquad (7\text{-}14)$$

$$+\frac{n^2\sigma_o^2}{(n\sigma_o^2+2\sigma^2)^2}\ A^2(A-t)^2\left[\frac{\sigma^2+\sigma_\epsilon^2}{n}+\frac{\sigma_o^2+\sigma_b^2+\sigma^2}{(n-1)(N-1)}\right] \qquad \text{(iii)}$$

The variance has three parts; the third part goes to zero as n goes to infinity i.e. variance of u and u_o goes to zero with increased replications. Likewise the second part goes to zero with increasing 'n' but, the first part

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

goes to $\frac{A^2(t-At)}{N-1}$. This can be easily seen by noting that as n increases to infinity 't' and consequently 'A' approach a limit. Thus if the number of varieties is very small the variance will be fairly high, though the replications may be many.

IV.2. Exponential

As worked out in section III the standard gain for the exponential is

$$e^{\frac{1}{2}(\lambda E^2-2\lambda\bar{a})}\left[\frac{\lambda E}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2} + \lambda c\ I(-c/E)\right]\frac{1}{P} \quad c=(\bar{a}-\lambda E^2)$$

and P is the fraction selected, it has been defined earlier but with the above substitution it will be

$$P=F(-\frac{\bar{a}}{E}) + e^{\frac{1}{2}(\lambda^2E^2-2\lambda\bar{a})}I(\frac{-c}{E})$$

Substituting from the analysis of variance table for λ and E the standard gain will take the form

$$e^{\frac{1}{2}(\frac{2f^2}{ny..^2}-\frac{2\bar{a}}{y..})}\left[\frac{f}{\sqrt{n\pi}.y..}\ \bar{e}^{c^2n/4f^2} + \frac{c}{y..}\ I(-c/f\sqrt{\frac{2}{n}})\right]\frac{1}{P} \qquad (7\text{-}25)$$

$$\text{and } P=F(-\bar{a}/\sqrt{\frac{2}{n}}.f\ ) + \text{Exp}(\frac{2f^2}{ny..^2} - \frac{2\bar{a}}{y..})I(-c/\sqrt{\frac{2}{n}}.f).\frac{1}{2}$$

Again we note that y.. and f are independent. For exponential the error is calculated on the assumption that u_o is known. That assumption does not effect the general principle, it just makes the formula a little shorter. Using the formula in (7-12) the variance of the standard gain can be worked out. The resulting expression for the variance is pretty long.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

It may be written as Var(standard Gain)$=A^2$Var(y..)$+B^2$Var(f). Variances for y.. and f have been given earlier.

$$A=e^{(\lambda^2E^2-2\lambda\bar{a})/2}\left[\left\{\frac{\lambda E}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}+\lambda c\ I(-c/E)(\bar{a}\lambda^2-\lambda^2E^2)-\frac{\lambda^2E}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}\right.\right.$$

$$\left.-\lambda^2cI(-c/E)+\lambda^2E^2I(-c/E)\right\}P-\left\{I(-c/E)(\lambda^2\bar{a}-\lambda^3E^2)+\bar{e}^{c^2/2E^2}.E\lambda^2\right\}$$

$$\left.\text{X Numerator}\right]\frac{1}{P^2}$$

Numerator is the numerator of the gain equation.

$$B=e^{\frac{1}{2}(\lambda^2E^2-2\lambda\bar{a})}\left[\left\{\frac{\lambda E}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}+c\ I(-c/E)\right\}E\lambda\sqrt{\frac{2}{n}}+\frac{\lambda}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}\right.$$

$$+\frac{\lambda}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}\left(\frac{c.\lambda.}{E}\frac{2\sqrt{2}}{\sqrt{n}}+\frac{c^2\sqrt{2}}{E^3\sqrt{n}}\right)-\sqrt{\frac{2}{n}}\frac{E^2\lambda^2}{\sqrt{2\pi}}I\left(\frac{-c}{E}\right)+\frac{c\lambda}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}\left(\frac{-4\lambda}{\sqrt{2n}}\frac{\sqrt{2}c}{E^2\sqrt{n}}\right)\left.\right]\frac{1}{P}$$

$$-\left[\frac{1}{\sqrt{2\pi}}\bar{e}^{\bar{a}^2/2E^2}\frac{.\bar{a}}{E^2}.\sqrt{\frac{2}{n}}+e^{\frac{1}{2}(\lambda^2E^2-2\bar{a}\lambda)}I(-c/E)\lambda^2E.\sqrt{\frac{2}{n}}\right.$$

$$\left.+e^{\frac{1}{2}(\lambda^2E^2-2\lambda\bar{a})}\frac{1}{\sqrt{2\pi}}\bar{e}^{c^2/2E^2}\left(\frac{-4\lambda}{\sqrt{2n}}-\frac{\sqrt{2}.c}{\sqrt{n}.E^2}\right)\right]\frac{\text{Numerator}}{P^2}$$

The practical utility of the above formula is doubtful. However it gives a striking difference from the normal distribution. In the case of exponential with the increase of n (replications) the standard error of gain goes to zero. This can be seen by noting that as n increases to infinity $\exp(\lambda^2E^2-2\lambda\bar{a})\frac{1}{2}$ has the limit $\bar{e}^{\lambda\bar{a}}$, $\bar{e}^{\bar{a}^2/2E^2}$ approaches zero and $\bar{e}^{c^2/2E^2}$ approaches zero. The difference from the normal distribution lies in the fact that the mean and variance of the exponential are not independent where as in the normal distribution they are. However in the exponential as well $(1/(N-1))$ is a factor for the variance of the standard

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gain, which leads to the same conclusion that if the number of varieties tested is small the error will be high.

V. Increased number of replications for the standard.

On intuitive basis it is reasonable to assume that precision can be increased by replicating the control more than other varieties; that will give a better idea for the performance of the control. We evaluate the consequences of having the replications for the control k times of the others. $\bar{X}_i$ will be based on n replications and $\bar{X}_o$ will be based on kn replications and we get

$$\text{Prob}(V_i \text{ with mean y selected}) = P(\bar{X}_i - \bar{X}_o > \delta a)$$

$$= I\left(\frac{\bar{a}-y}{\sigma}\sqrt{\frac{k\,n}{k+1}}\right)$$

and from that we get the fraction selected for the normal distribution.

$$\text{Fraction selected} = \frac{1}{\sqrt{2\pi}\sigma_o}\int_{-\infty}^{\infty} e^{-\frac{(y-u)^2}{2\sigma_o^2}} \int_{\frac{(\bar{a}-y)}{\sigma}\sqrt{\frac{kn}{k+1}}}^{\infty} e^{-t^2/2}\, dt\, dy \qquad (7\text{-}26)$$

Substitute $\frac{(y-u)}{\sigma_o}\sqrt{\frac{kn}{k+1}} = z_2$ and $\frac{(X-u+u_o)}{\sigma}\sqrt{\frac{kn}{k+1}} = z_1$ then (7-26) reduces to

$$\int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}\sigma_{z_1}\sigma_{z_2}} \exp\left(\frac{Z_1^2}{\sigma_{z_1}^2} + \frac{Z_2^2}{\sigma_{z_2}^2} - \frac{2\rho Z_1 Z_2}{\sigma_{z_1}\sigma_{z_2}}\right) \frac{-1}{2(1-\rho^2)}\, dZ_1\, dZ_2 \qquad (7\text{-}27)$$

This is a bivariate integral with variance covariance matrix

Z_1	Z_2
$1 + \frac{k\,m^2 n}{k+1}$	$\frac{k.m.n}{k+1}$
$\frac{kmn}{k+1}$	$\frac{kn}{k+1}$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Change to standardized variables and integrate out Z_2 then it becomes $\frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-Z_1^2/2} dZ_1$ where $t = \frac{\bar{a}-u}{\sqrt{E'^2-\sigma_0^2}}$ and $E' = \sqrt{\frac{(k+1)}{kn}} \cdot \sigma$. Using the same type of arguments as in section II for deriving (7-15) the gain in the yielding capacity comes to be

$$\frac{\sigma_0}{\sqrt{E'^2 + \sigma_0^2}} \quad \frac{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{\frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-t^2/2} dt}$$

This is of the same form as (7-15), the only difference is that E has been replaced by E' and E' is a fraction of E. Thus the effect of increasing the replications for the standard is the reduction of E or equivalently of σ_0^2 the error variance; so the precision can be increased by replicating the standard more than the others. This device will be particularly effective if the number of varieties is so large that they cannot be replicated many times.

VI. Higher moments of the population after first stage selection.

(a) For the exponential case we may write I_n' (nth moment) as $I_n' = \int_0^\infty \lambda y^n e^{\lambda y} F(\frac{y-\bar{a}}{E}) dy / P$, P is fixed and may be ignored and write $I_n = \int_0^\infty \lambda y^n e^{\lambda y} F(\frac{y-\bar{a}}{E}) dy$. To determine I_n recursive relation can be developed; integrating by parts

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$I_n = n\int_0^\infty e^{-\lambda y} y^{n-1} F\left(\frac{y-\bar{a}}{E}\right)dy + \frac{1}{E}\int_0^\infty e^{-\lambda y} y^n f\left(\frac{y-\bar{a}}{E}\right) dy$$

$$= nI_{n-1} + t_n \exp(\lambda^2E^2 - 2\lambda\bar{a})\ \frac{1}{2}$$

$$t_n = \frac{1}{E}\int_0^\infty e^{-\lambda y} y^n f\left(\frac{y-\bar{a}}{E}\right)dy$$

$$= (n-1)E^2\ t_{n-2} + (\bar{a}-\lambda E^2)\ t_{n-1}$$

(b) For the normal case only second moment for $Z_2=\frac{(y-u)}{\sigma_o}$ is evaluated

$$E(Z_2-\rho Z_1)^2 = E(Z_2^2 - 2\rho Z_1Z_2 + Z_1^2) = \frac{1}{2\pi P(1-\rho^2)}\int_{-\infty}^{\infty}\int_t^\infty (Z_2-\rho Z_1)^2 e^{\frac{-(Z_1^2+Z_2^2-2\rho Z_1Z_2)}{(1-\rho^2)2}} dz_1 dz_2$$

Changing the order of integration and integrating Z_2 we get

$$\frac{(1-\rho^2)}{P\sqrt{2\pi}}\int_t^\infty e^{-t^2/2}\ dt = 1-\rho^2 \qquad (7\ 28)$$

$$E(Z_1-\rho Z_2)^2 = \frac{1}{2\pi(1-\rho^2)^{1/2}P}\int_{-\infty}^{\infty}\int_t^\infty (Z_1-\rho Z_2)^2\ e^{-\frac{1}{2}\frac{(Z_1^2+Z_2^2-2\rho Z_1Z_2)}{(1-\rho^2)}}\ dZ_1\ dZ_2$$

Integrating the inner integral we get

$$\frac{1-\rho^2}{2\pi(1-\rho^2)^{\frac{1}{2}}}\ \frac{1}{P}\int_{-\infty}^{\infty}\left\{(Z_1-\rho Z_2)e^{\frac{-(Z_1^2+Z_2^2-2\rho Z_1Z_2)}{2(1-\rho^2)}}\Big|_t^\infty + \int_t^\infty e^{\frac{(Z_1^2+Z_2^2-2\rho Z_1Z_2)}{(1-\rho^2)2}} dZ_1\right\} dZ_2$$

$$= (A) + (B)$$

Solving these separately we get

$$(A) = \lambda(t)(1-\rho^2)\ t.$$

where, $\lambda(t) = \frac{1}{\sqrt{2\pi}}\ \frac{e^{\frac{-t^2}{2}}}{I(t)}$

$(B) = 1-\rho^2$ and from that we get

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$E(Z_1^2+Z_2^2-2\rho Z_1 Z_2)=(A)+(B)=(1-\rho^2)\left[\lambda(t).t(1-\rho^2)+1\right] \qquad (7\text{-}29)$$

Subtracting (7-28) from (7-29) we have

$$(1-\rho^2)\left[E(Z_1^2-Z_2^2)\right] = (1-\rho^2)^2\lambda(t).t$$

$$\text{or } E(Z_2^2) = E(Z_1^2)-(1-\rho^2)\lambda(t).t$$

Z_1 is a standardized normal variable and the second moment of a standardized normal variable after truncation at 't' has been given by Cramer (62), $t.\lambda(t)+1$. Using this result we get the second moment of Z_2 as

$$E(Z_2^2) = \rho^2\lambda(t)t + 1$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 8

SECOND STAGE SCREENING

At the second stage there are two possible procedures to follow. Firstly, the selection at the second stage is based on the results of second stage observations only and secondly, the results of the first stage are also taken into consideration. In the first three sections consequences of the selection based upon second stage results only are evaluated.

8.1 Without considering the first stage result

I. Normal

At the second stage the procedure would be to select a variety if it exceeds the standard by 'bδ', 'b' may be the same as 'a' used for the first stage but, it is reasonable to suppose that in general 'b' will be greater than 'a'. Thus variety V_i will be selected at the second stage if $\bar{\bar{X}}_i - \bar{\bar{X}}_o > b\delta$, n_1 and n_2 denote the number of replications for the first and second stage respectively.

Prob.(V_i with mean Y selected in two stage) = $P(\bar{\bar{X}}_i - \bar{\bar{X}}_o > b\delta,\ \bar{X}_i - \bar{X}_o > a\delta)$, as $\bar{\bar{X}}_i - \bar{\bar{X}}_o$ and $\bar{X}_i - \bar{X}_o$ are independent therefore the above stated probability will be the product of two terms i.e.

$$\text{Prob}(V_i \text{ is selected}) = P(\bar{\bar{X}}_i - \bar{\bar{X}}_o > \delta b).P(\bar{X}_i - \bar{X}_o > \delta a)$$

$$= I\left(\frac{\bar{a}-Y}{\sqrt{2/n_1}\sigma}\right) I\left[(\bar{b}-Y)/\sqrt{2/n_2}\sigma\right]$$

With this the fraction selected at the second stage will be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{\sqrt{(n_1 n_2)}}{(2\pi)^{3/2} 2\sigma_o \sigma^2} \int_{-\infty}^{\infty} \int_{b\delta}^{\infty} \int_{a\delta}^{\infty} \text{Exp}-\tfrac{1}{2}\left[\frac{(Y-u)^2}{\sigma_o^2} + \frac{(X_1-Y+u_o)^2}{\sigma^2 \frac{2}{n_1}} + \frac{(X_2-Y+u_o)^2}{\sigma^2 \frac{2}{n_2}} \right] dX_1 dX_2 dY$$

(8-1)

let $n_2 = \frac{n_1}{W}$ and $w = \sqrt{W}$ W is a positive real number.

In (8-1) substitute $\sqrt{n_2/2} \cdot \left(\frac{Y-u}{\sigma_o}\right) = Z_3$

$$\sqrt{n_1/2} \cdot \left(\frac{X_1-u+u_o}{\sigma}\right) = Z_1$$

and $$\sqrt{n_2/2}\left(\frac{X_2-u+u_o}{\sigma}\right) = Z_2$$

With these substitutions (8-1) becomes

$$\frac{1}{(2\pi)^{3/2}} \sqrt{\frac{2w^2}{n_1}} \int_{-\infty}^{t_2'} \int_{t_1'}^{\infty} \int \text{Exp}-\tfrac{1}{2}\left[\left(\frac{2}{n_2} + w^2 \frac{\sigma_o^2}{\sigma^2} + \frac{\sigma_o^2}{\sigma^2}\right) Z_3^2 - 2Z_1 Z_3 \frac{\sigma_o}{\sigma} + Z_1^2 + Z_2^2 - 2\frac{\sigma_o}{\sigma} Z_2 Z_3 \right] dZ_1 dZ_2 dZ_3$$

(8-2)

$$t_1' = \left(\frac{\bar{a}-u}{\sigma}\right)\sqrt{\frac{n_1}{2}} \quad \text{and} \quad t_2' = \left(\frac{\bar{b}-u}{\sigma}\right)\sqrt{\frac{n_2}{2}}$$

This is a trivariate normal integral with variance covariance matrix

Z_1	Z_2	Z_3
$1+n_2\frac{w^2m^2}{2}$	$n_2\frac{w\ m^2}{2}$	$n_2\ \frac{w\ m}{2}$
	$1+n_2\ \frac{m^2}{2}$	$n_2\ \frac{m}{2}$
		$\frac{n_2}{2}$

$m = \frac{\sigma_o}{\sigma}$

Again transforming to standardized variables (8-2) transforms to

$$\frac{1}{(2\pi)^{3/2}|A|^{1/2}} \int_{-\infty}^{\infty} \int_{t_2}^{\infty} \int_{t_1}^{\infty} f(Z_1 Z_2 Z_3)\ dZ_1 dZ_2 dZ_3$$

$$t_2 = \left(\frac{\bar{b}-u}{\sigma}\right)\sqrt{n_2/(n_2m^2+2)}$$

$$t_1 = \left(\frac{\bar{a}-u}{\sigma}\right)\sqrt{n_1/(2+n_2w^2m^2)}$$

'A" is the determinant of the correlation matrix determined from the above given variance covariance matrix and 'f' is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a trivariate normal density function with the same correlation structure. Integrating out Z_3 will give a bivariate normal integral with correlation given by the top left hand corner submatrix. Thus the fraction selected at the second stage will be

$$\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_{t_2}^{\infty}\int_{t_1}^{\infty} \text{Exp}-\tfrac{1}{2}(Z_1^2+Z_2^2-2\rho Z_1Z_2)\frac{1}{1-\rho^2}\, dZ_1dZ_2 \qquad (8\text{-}3)$$

t_1 and t_2 have already been defined and $\rho = \dfrac{n_2wm^2}{\sqrt{(2+n_2m^2)(2+n_2w^2m^2)}}$

The mean of Z_3 i.e. of $\frac{(Y-u)}{\sigma_o}$, standard gain depends upon Z_1 and Z_2; Z_3 has a regression on Z_1 and Z_2; thus, the mean of Z_3 may be written as $E(Z_3)=\beta_2E(Z_2)+\beta_1E(Z_1)$; where β_1 and β_2 are the partial regressions of Z_3 on Z_1 and Z_2 respectively. Because Z_1, Z_2 and Z_3 are standardized variables the partial regressions will be the partial correlations. Therefore to obtain the standard gain we evaluate $E(Z_2-\rho Z_1)$ and $E(Z_1-\rho Z_2)$

$$E(Z_2-\rho Z_1) = \frac{1}{P.2\pi(1-\rho^2)^{\frac{1}{2}}} \int_{t_2}^{\infty}\int_{t_1}^{\infty}\left\{\exp-\frac{1}{2(1-\rho^2)}(Z_1^2+Z_2^2-2\rho Z_1Z_2)\right\}(Z_2-\rho Z_1)\,dz_1dz_2$$

$$= \frac{(1-\rho^2)}{P.(2\pi)(1-\rho^2)^{\frac{1}{2}}} \int_{t_1}^{\infty} \exp-\tfrac{1}{2}\frac{(t_2^2+Z_1^2-2\rho t_2Z_1)}{(1-\rho^2)}\, dZ_1$$

$$= \frac{1-\rho^2}{P\sqrt{2\pi}} \cdot e^{-t_2^2/2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{\frac{t_1-\rho t_2}{\sqrt{1-\rho^2}}}^{\infty} e^{-y^2/2}\, dy = \frac{1-\rho^2}{P}\,\frac{e^{-t_2^2/2}}{\sqrt{2\pi}} I(t_2) \qquad (8\text{-}4)$$

$$\text{Similarly } E(Z_1-\rho Z_2) = \frac{1-\rho^2}{P}\,\frac{e^{-t_1^2/2}}{\sqrt{2\pi}}\, I(t_1) \qquad (8\text{-}5)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From (8-4) and (8-5) we get

$$E(Z_1) = \frac{1}{B\sqrt{2\pi}} \left[e^{-t_1^2/2} I(t_1) + \rho e^{-t_2^2/2} I(t_2) \right]$$

$$E(Z_2) = \frac{1}{B\sqrt{2\pi}} \left[\rho e^{-t_1^2/2} I(t_1) + e^{-t_2^2/2} I(t_2) \right] \qquad (8\text{-}6)$$

$E(Z_3) = \beta_1 E(Z_1) + \beta_2 E(Z_2)$; β_1 and β_2 can be determined from the covariance matrix

II. Exponential with parameter λ

As in section I $\Pr(V_i \text{ selected}) = F\left(\frac{Y-\bar{a}}{E}\right) F\left(\frac{Y-\bar{b}}{E_2}\right)$ Where $E = \sqrt{2/n_1}\cdot\sigma$ and $E_2 = \sqrt{2/n_2}\cdot\sigma$. That will give the fraction selected as

$$= \int_0^\infty \lambda e^{-\lambda Y} F\left(\frac{Y-\bar{a}}{E}\right) F\left(\frac{Y-\bar{b}}{E_2}\right) dY$$

$$= F\left(\frac{-\bar{a}}{E}\right) F\left(\frac{-\bar{b}}{E_2}\right) + \frac{1}{E_2} \int_0^\infty e^{-\lambda Y} F\left(\frac{Y-\bar{a}}{E}\right) f\left(\frac{Y-\bar{b}}{E_2}\right) dY$$

$$+ \frac{1}{E} \int_0^\infty e^{-\lambda Y} f\left(\frac{Y-\bar{a}}{E}\right) F\left(\frac{Y-\bar{b}}{E_2}\right) dY \qquad (8\text{-}7)$$

The first term is a product of two univariate integrals. The second and the third terms are bivariate integrals. Solution of the second term is indicated below and that of the third can be written by noting the symmetry.

Second term is $\frac{1}{2\pi E E_2} \int_0^\infty \int_0^\infty e^{-\lambda Y} e^{-(Y-\bar{b})^2/2E_2^2} \cdot e^{-\frac{(X+\bar{a}-Y)^2}{2E^2}} dx\,dy$

$$= \frac{1}{2\pi E E_2} \int_0^\infty \int_0^\infty \text{Exp} -\tfrac{1}{2}\left[Y^2\left(\frac{1}{E^2} + \frac{1}{E_2^2}\right) + \frac{X^2}{E^2} - 2Y\left(\frac{\bar{b}}{E_2^2} + \frac{\bar{a}}{E^2} - \lambda\right) + \frac{2\bar{a}X}{E^2} + \frac{\bar{b}^2}{E_2^2} + \frac{\bar{a}^2}{E^2} \right] dx\,dy$$

$$\times e^{\left(\frac{E_1^2\lambda^2 - 2\lambda\bar{b}}{2}\right)} \qquad (8\text{-}8)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

It is bivariate integral with variances and correlation as follows

$$\text{Var}(X) = E^2 + E_2^2 \qquad \text{mean of } X = \bar{b}-\bar{a}-E_2^2\lambda$$

$$\text{Var}(Y) = E_2^2 \qquad \text{mean of } Y = \bar{b}-E_2^2\lambda$$

$$\text{Cov}(XY) = E_2^2$$

Changing to standardized variables (8-8) becomes

$$\int_{t_2}^{\infty}\int_{t_1}^{\infty} f_2\rho\,(Z_1Z_2)\,dZ_1dZ_2.\exp(E_2^2\lambda^2-2\lambda\bar{b})/2 \qquad (8\text{-}9)$$

where $t_1 = -\dfrac{(\bar{b}-\bar{a}-E_2^2\lambda)}{\sqrt{E^2+E_2^2}}$ and $t_2 = -\dfrac{(b-E_2^2\lambda)}{E_2}$

Noting the symmetry, third part of (8-7) can be written as

$$\int_{t_2}^{\infty}\int_{t_1}^{\infty} f_2\rho'(Z_1Z_2)dZ_1dZ_2\exp(E^2\lambda^2-2\lambda\bar{a})/2 \qquad (8\text{-}10)$$

where $t_1' = -\dfrac{(\bar{a}-\bar{b}-E^2\lambda)}{\sqrt{E^2+E_2^2}}$ and $t_2' = -\dfrac{(\bar{a}-E^2\lambda)}{E}$ and $\rho' = \dfrac{E}{\sqrt{E^2+E_2^2}}$

For the gain after second stage selection again we work in terms of standard gain, i.e., in terms of $\lambda(Y-1)$

$$E(\lambda Y-1) = \int_0^{\infty} Y\lambda^2\,\bar{e}^{\lambda Y}F\left(\frac{Y-\bar{a}}{E}\right)F\left(\frac{Y-\bar{b}}{E_2}\right)dY/P - 1$$

$$= \int_0^{\infty}\lambda\bar{e}^{\lambda Y}F\left(\frac{Y-\bar{a}}{E}\right)F\left(\frac{Y-\bar{b}}{E_2}\right)dY/P + \frac{1}{P\,E_2}\,\lambda\int_0^{\infty}\bar{e}^{\lambda Y}YF\left(\frac{Y-\bar{a}}{E}\right)f\left(\frac{Y-\bar{b}}{E_2}\right)dY + \frac{1}{PE}\,\lambda\int_0^{\infty}\bar{e}^{\lambda Y}Yf\left(\frac{Y-\bar{a}}{E}\right)dY$$

$$-1$$

The numerator in the first term is the same as (8-7) i.e. fraction 'P' selected thus it will cancel with '1' at the end and we are left with

$$\frac{1}{P\,E_2}\int_0^{\infty}\lambda\bar{e}^{\lambda Y}Y\,F\left(\frac{Y-\bar{a}}{E}\right)f\left(\frac{Y-\bar{b}}{E_2}\right)dY + \frac{1}{PE}\int_0^{\infty}\lambda\bar{e}^{\lambda Y}Yf\left(\frac{Y-\bar{a}}{E}\right)F\left(\frac{Y-\bar{b}}{E_2}\right)dY \qquad (8\ 11)$$

We solve these two terms separately. Compare the first term

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of (8-11) - call it A_1 - with the second term of (8-7) - call it A - Ignoring the constant P, A_1 is the expectation of Y with respect to A. Thus making similar transformations as used in the solution of 'A' and noting its solution given in (8-9) we can write 'A_1' as

$$\lambda\left[\frac{1}{P(2\pi)(1-\rho^2)^{1/2}}\int_{t_2}^{\infty}\int_{t_1}^{\infty}E_2Z_2e^{-\frac{1}{(1-\rho^2)}(Z_1^2+Z_2^2-2\rho Z_1Z_2)}dZ_1dZ_2\cdot e^{(E_2^2\lambda^2-2\lambda\bar{b})/2}\right]$$

$+\lambda(A)(\bar{b}-E_2^2\lambda)/P$, t_1, t_2 and ρ have been defined in the solution of 'A'.

Similarly compare second part of (8-11) - call it B_1 - with third term of (8-7) - call it B - Noting the solution of B in (8-10) we can write B_1

$$\frac{\lambda}{P}\left[e^{\frac{(\lambda^2E_2^2+2\bar{a}\lambda)}{2}}\frac{2\pi}{2\pi}\int_{t_2'}^{\infty}\int_{t_1'}^{\infty}EZ_2f_{2,\rho'}(Z_1Z_2)dZ_1dZ_2\right]-(B)(\bar{a}-E^2\lambda)\frac{\lambda}{P}$$

t_1', t_2' and ρ' have been defined for the solution of 'B'.

Thus in principle the problem reduces to find the value of the integral of the form

$$\frac{2\pi}{2\pi}\int_{t_2}^{\infty}\int_{t_1}^{\infty}Z_2f_{2,\rho'}(Z_1\ Z_2)dZ_1\ dZ_2$$

This type of integral has been evaluated in Section I.1, equation (8-6). Using that result it may be written as

$$\frac{1}{\sqrt{2\pi}}\left[\rho e^{-t_1^2/2}I(t_1)+e^{-t_2^2/2}I(t_2)\right] \text{ where } I(t_1)=\frac{1}{\sqrt{2\pi}}\int_{\frac{t_2-et_1}{\sqrt{1-e^2}}}^{\infty}e^{-y^2/2}dy,$$

$I(t_2)$ is defined analogously.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

III. Double Exponential

In this section results are obtained on the assumption that the varietal means follow double exponential distribution with parameters λ and u and the results of the 1st stage are not taken into consideration. Double exponential is characterized by the density function

$$\tfrac{1}{2}\,\lambda\, e^{-\lambda|Y-u|} \qquad -\infty < Y < \infty$$

As before Prob(that V_i with mean Y selected) $= F\left(\frac{Y-\bar{a}}{E}\right) F\left(\frac{Y-b}{E_2}\right)$

That will give the fraction of the population selected at the second stage.

$$\text{Fraction selected} = \tfrac{1}{2}\int_{-\infty}^{\infty} \lambda e^{-\lambda|Y-u|} F\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right)\, dY = P$$

$$= \tfrac{1}{2}\left[e^{\lambda u}\int_{u}^{\infty} e^{-\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dy + e^{-\lambda u}\int_{-\infty}^{u} e^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dy \right]$$

$$= \tfrac{1}{2}\,(A + B)$$

$$(A) = F\left(\frac{u-\bar{b}}{E_2}\right) F\left(\frac{u-\bar{a}}{E}\right) + \frac{1}{E}\, e^{\lambda u}\int_{u}^{\infty} e^{-\lambda Y}\, f\left(\frac{Y-\bar{a}}{E}\right) F\left(\frac{Y-\bar{b}}{E_2}\right) dY$$

$$+ \frac{1}{E_2}\, e^{\lambda u}\int_{u}^{\infty} e^{-\lambda Y} f\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dY \qquad (8\text{-}12)$$

$$(B) = F\left(\frac{u-\bar{b}}{E_1}\right) F\left(\frac{u-\bar{a}}{E}\right) - \frac{1}{E} e^{-\lambda u}\int_{-\infty}^{u} e^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) f\left(\frac{Y-\bar{a}}{E}\right) dY$$

$$- \frac{1}{E_2} e^{-\lambda u}\int_{-\infty}^{u} e^{\lambda Y} f\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dY \qquad (8\text{-}13)$$

First part of (A) is the product of univariate integrals and the second and the third parts are bivariate integrals but they require to be transformed into proper form.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\text{Second part of (A)} = \frac{1}{E} e^{\lambda u} \int_u^\infty e^{-\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) f\left(\frac{Y-\bar{a}}{E}\right) dy$$

$$= \frac{1}{EE_2} \frac{1}{2\pi} e^{\lambda u} \int_u^\infty \int_o^\infty e^{-\lambda Y} e^{-\frac{(Y-\bar{a})^2}{2E^2} - \frac{(X-\bar{b}-Y)^2}{2E_2^2}} dX dY$$

$$= \int_{t_2}^\infty \int_{t_1}^\infty f_{2,\rho}(x,y) dx dy \; e^{\frac{1}{2}(\lambda^2 E^2) - \lambda(\bar{a}-u)} \qquad (8\text{-}14)$$

when $\rho = \dfrac{E}{\sqrt{E^2+E_2^2}}$, $t_1 = \dfrac{-(\bar{a}-\bar{b}-\lambda E^2)}{\sqrt{E_1^2+E_2^2}}$

$$t_2 = \frac{(u-\bar{a}+\lambda E^2)}{E}$$

Third part of (A) can be written down by observing the symmetry and it is

$$\int_{t_2'}^\infty \int_{t_1'}^\infty f_{2\rho'}(xy)_{dx\ dy} \; e^{\frac{1}{2}\lambda^2 E_2^2 - \lambda(\bar{b}-u)} \qquad (8\text{-}15)$$

$$t_1' = -\frac{(\bar{b}-\bar{a}-\lambda E_2^2)}{\sqrt{E_1^2+E_2^2}}$$

$$t_2' = \frac{u-\bar{b}+\lambda E_2^2}{E_2}$$

$$\rho' = \frac{E_2}{\sqrt{E_1^2+E_2^2}}$$

First part of (B) is a product of univariate integrals, second and third parts of (B) can be transformed into bivariate form.

$$\text{Second part of (B)} = \frac{e^{-\lambda u}}{EE_2} \int_{-\infty}^u e^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) f\left(\frac{Y-\bar{a}}{E}\right) dY$$

$$= \frac{e^{-\lambda u}}{EE_2 2\pi} \int_{-\infty}^u \int_{-\infty}^o e^{\lambda Y} e^{-\frac{1}{2}\left(\frac{Y-\bar{a}}{E}\right)^2 - \frac{1}{2}\left(\frac{X+Y-\bar{b}}{E_2}\right)^2} dX dY$$

$$= \int_{-\infty}^{\ell_2} \int_{-\infty}^{\ell_1} f_{\underline{2}R}(x,y) dx\ dy \; e^{\frac{\lambda^2 E^2}{2} + \lambda(\bar{a}-u)} \qquad (8\text{-}16)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\ell_1 = \frac{-(\bar{b}-\bar{a}-\lambda E_1^2)}{\sqrt{E^2+E_2^2}}$$

$$\ell_2 = \frac{(u-\bar{a}-\lambda E^2)}{E}$$

$$R = \frac{-E}{\sqrt{E+E_2^2}}$$

By noting the symmetry the third part of (B) can be written down as

$$\int_{-\infty}^{\ell_2'}\int_{-\infty}^{\ell_1'} f_{2,R'}(x,y)dx\,dy\; e^{\frac{\lambda^2E_2^2}{2}+\lambda(\bar{a}-u)} \qquad (8\text{-}17)$$

$$\ell_1' = \frac{-(\bar{a}-\bar{b}-\lambda E_2^2)}{\sqrt{E+E_2^2}}$$

$$\ell_2' = \frac{(u-\bar{b}-\lambda E_2^2)}{E_2}$$

$$R' = \frac{-E_2}{\sqrt{E_1^2+E_2^2}}$$

Combining all these we may write the fraction selected as

$$P=\tfrac{1}{2}\Big|\, 2F\Big(\frac{u-\bar{b}}{E_2}\Big)F\Big(\frac{u-\bar{a}}{E}\Big)+e^{\frac{\lambda^2E_1^2}{2}-\lambda(\bar{a}-u)}I_{2,\rho}(t_1t_2)+e^{\frac{\lambda^2E_2^2}{2}-\lambda(\bar{b}-u)}I_{2,\rho'}(t_1't_2')$$
$$-e^{\frac{\lambda^2E^2}{2}+\lambda(\bar{a}-u)}F_{2,R}(\ell_1,\ell_2)-F_{2,R'}(\ell_1',\ell_2')e^{\frac{\lambda^2E_2^2}{2}+\lambda(\bar{b}-u)} \qquad (8\text{-}18)$$

Variance in double exponential is $2/\lambda^2$, that will give the standard gain as $\frac{E(Y-u)}{\sqrt{2}/\lambda}$

$$= \frac{1}{P\;2\sqrt{2}}\int_{-\infty}^{\infty}\lambda^2 Y e^{-\lambda|Y-u|}F\Big(\frac{Y-\bar{b}}{E_2}\Big)F\Big(\frac{Y-\bar{a}}{E}\Big)dY - \frac{\lambda u}{\sqrt{2}}$$

$$= \frac{1}{2\sqrt{2}.P}\,e^{\lambda u}\int_{u}^{\infty}\lambda^2 Y\, e^{-\lambda Y}F\Big(\frac{Y-\bar{b}}{E_2}\Big)F\Big(\frac{Y-\bar{a}}{E}\Big)dy + \frac{1}{P.2\sqrt{2}}\int_{-\infty}^{u}\lambda^2\, Y\, e^{-\lambda u}e^{\lambda Y}F\Big(\frac{Y-\bar{b}}{E_2}\Big)F\Big(\frac{Y-\bar{a}}{E}\Big)dy$$
$$- \frac{\lambda u}{\sqrt{2}}$$

$$= \frac{1}{P}\Big\{(C)+(D)\Big\} - \frac{\lambda u}{\sqrt{2}}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$(C) = \frac{1}{2\sqrt{2}} \Big| \lambda u . F\left(\frac{u-\bar{b}}{E_2}\right) F\left(\frac{u-\bar{a}}{E}\right) + \frac{\lambda}{E} e^{\lambda u} \int_u^{\infty} Y . \bar{e}^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) f\left(\frac{Y-\bar{a}}{E}\right) dY$$

$$+ \frac{\lambda}{E_2} \int_u^{\infty} Y . \bar{e}^{\lambda Y} f\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dY + \lambda\, e^{\lambda u} \int_u^{\infty} F\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) \bar{e}^{\lambda Y} dy \Big]$$

$$(D) = \frac{\lambda \bar{e}^{\lambda u}}{2\sqrt{2}} \Big| u F\left(\frac{u-\bar{b}}{E_2}\right) F\left(\frac{u-\bar{a}}{E}\right) e^{\lambda u} - \frac{1}{E} \int_{-\infty}^{u} Y\, e^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) f\left(\frac{Y-\bar{a}}{E}\right) dy$$

$$- \frac{1}{E_2} \int_{-\infty}^{u} Y\, e^{\lambda Y} f\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dy - \int_{-\infty}^{u} e^{\lambda Y} F\left(\frac{Y-\bar{b}}{E_2}\right) F\left(\frac{Y-\bar{a}}{E}\right) dy \Big]$$

Solution of such integrals has been given in section II, so it will be sufficient to give the final result without going through lengthy algebra.

$$\text{Standard gain} = \frac{1}{2\sqrt{2}.P} \Big[2\lambda u F\left(\frac{u-\bar{b}}{E_2}\right) F\left(\frac{u-\bar{a}}{E}\right) + e^{\frac{\lambda^2 E^2}{2} - \lambda(\bar{a}-u)} I_{2,\rho}(t_1 t_2)$$

$$+ e^{\frac{\lambda^2 E_2^2}{2} - \lambda(\bar{b}-u)} I_{2,\rho'}(t_1', t_2') + e^{\frac{\lambda^2 E^2}{2} - \lambda(\bar{a}-u)} \Big\{ \lambda E I_{2,\rho}(t_1 t_2 - Y) + \lambda(\bar{a}-\lambda E^2) I_{2,\rho}(t_1 t_2) \Big\}$$

$$+ e^{\frac{\lambda^2 E_2^2}{2} - \lambda(\bar{b}-u)} \Big\{ \lambda E_2 I_{2\rho'}(t_1' t_2' - Y) + \lambda(\bar{b}-\lambda E_2^2) I_{2,\rho'}(t_1' t_2') \Big\}$$

$$+ e^{\frac{\lambda^2 E^2}{2} + \lambda(\bar{a}-u)} F_{2,R}(\ell_1, \ell_2) + e^{\frac{\lambda^2 E_2^2}{2} + \lambda(\bar{b}-u)} F_{2,R'}(\ell_1', \ell_2')$$

$$- e^{\frac{\lambda^2 E^2}{2} + \lambda(\bar{a}-u)} \Big\{ \lambda E\, F_{2R}(\ell_1, \ell_2 - Y) + \lambda(a+\lambda E^2) F_{2R}(\ell_1, \ell_2) \Big\}$$

$$- e^{\frac{\lambda^2 E_2^2}{2} + \lambda(\bar{b}-u)} \Big\{ \lambda E_2 F_{2,R'}(\ell_1' \ell_2' - Y) + \lambda(\bar{b}+\lambda E_2^2) F_{2,R'}(\ell_1', \ell_2') \Big\} \Big] - \frac{\lambda u}{\sqrt{2}} \qquad (8\text{-}19)$$

Where $F_{2,R'}(\beta_1', \beta_2' - Y) = \int_{-\infty}^{\ell_2'} \int_{Y}^{\ell_1'} f_{2R'}(x, y)\, dx\, dy$

and $I_{2,\rho}(t_1, t_2 - Y) = \int_{t_2}^{-\infty} \int_{Y}^{-\infty} f_{2\rho}(x, y)\, dx\, dy$

The constants such as t_1, t_2, etc. have already been defined in the evaluation of P the fraction selected.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

IV. Normal

Alternatively it may be desirable that at the second stage the result of the first stage is also considered. In the following sections consequences of this consideration are evaluated. Number of replications in the first stage $=n_1$. Number of replications in the second stage $n_2 = n_1/w^2$. Thus the selection at the second stage is based on $n_1+\frac{n_1}{w^2}=n_1\left(1+\frac{1}{w^2}\right)=n$ replications. Denoting $w^2/(1+w^2)$ by w'^2 we get $n=n_1/w'^2$. Selection at the first stage is based on $X_1=\bar{X}_1-\bar{X}_0$; and at the second stage the selection decision will depend upon $X_2=\bar{\bar{X}}_1-\bar{\bar{X}}_0$; X_2 is based upon n_1/w^2replications which include the replications conducted at the first stage. X_1 and X_2 are correlated with correlation $r=\sqrt{n_1/n} = w'$ and the variance of X_1 and X_2 are $\frac{2\sigma^2}{n_1}$ and $\frac{\sigma^2 2}{n}$ respectively. That will give

Prob.(A variety with mean Y selected)$=P(X_1 > \delta a, X_2 > \delta b)$

$$=I_{2,r}\left(\frac{\bar{a}-Y}{\sqrt{\frac{2}{n_1}}\cdot\sigma}, \frac{\bar{b}-Y}{\sqrt{\frac{2}{n}}\cdot\sigma}\right) \qquad (8\text{-}20)$$

That will give the fraction of the population selected as

$$\frac{1}{\sqrt{2\pi}\sigma_o}\int_{-\infty}^{\infty} e^{-(Y-u)^2/2\sigma_o^2} I_{2,r}\left(\frac{\bar{a}-Y}{\sqrt{\frac{2\sigma}{n_1}}}, \frac{\bar{b}-Y}{\sqrt{\frac{2}{n}}\sigma}\right) dY$$

$$= \frac{1}{(1-r^2)^{\frac{1}{2}}(2\pi)^{\frac{3}{2}}\sigma\sigma^2 2}\int_{-\infty}^{\infty}\int_{\delta b}^{\infty}\int_{a\delta}^{\infty} \exp -\frac{1}{2}\left|\frac{1}{1-r^2}\left(\frac{(X_1-Y+u_o)^2 n_1}{\sigma^2\ 2}+\frac{(X_2-Y+u_o)^2 n}{\sigma^2\ 2}\right.\right.$$

$$\left.\left.-\frac{2r(X-Y+u_o)}{\sigma^2}\frac{\sqrt{n_1 n}}{2} - \frac{(Y-u)^2}{\sigma_o^2}\right)\right| dX_1 dX_2 dY \qquad (8\text{-}21)$$

In (8-21) substitute $\sqrt{\frac{n}{2}}\left(\frac{Y-u}{\sigma_o}\right)=Z_3'$

$\sqrt{\frac{n_1}{2}}\cdot\left(\frac{X_1-u+u_o}{\sigma}\right) = Z_1'$, and $\sqrt{\frac{n_1}{2}}\cdot\left(\frac{X_2-u+u_o}{\sigma}\right) = Z_2'$,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

then it reduces to

$$\frac{1}{(1-r^2)^{\frac{1}{2}}(2\pi)^{\frac{3}{2}}\sqrt{n}} \int_{\infty}^{\infty}\int_{t_2'}^{\infty}\int_{t_1'}^{\infty} \exp -\frac{1}{2}\left[Z_3'^2\left\{\frac{2}{n}+\frac{1}{1-r^2}(w'^2m^2+m^2-2rw'm)\right\}+\frac{Z_1'^2}{1-r^2}+\frac{Z_2'^2}{1-r^2}\right.$$

$$\left.-2Z_1'Z_2'm\left(\frac{w'-r}{1-r^2}\right)-2\cdot Z_2'Z_3'm\left(\frac{1-rw'}{1-r^2}\right)\right] dZ_1'dZ_2'dZ_3' \qquad (8\text{-}22)$$

This is a trivariate normal integral with covariance matrix

Z_1	Z_2	Z_3
$1+m^2w'^2\ \frac{n}{2}$	$w'+m^2w'\ \frac{n}{2}$	$n\ w'\ \frac{n}{2}$
	$1+m^2.\frac{n}{2}$	$m.\frac{n}{2}$
		$\frac{n}{2}$

Changing to standardized variables and integrating out Z_3 (8-22) reduces to $\int_{t_2}^{\infty}\int_{t_1}^{\infty} f_{2,\rho}(Z_1Z_2)\,dZ_1\,dZ_2$, where $\rho = \frac{2w'+m^2.w'.n}{\sqrt{(2+m^2.n)(2+m^2w'^2n)}}$

$t_1 = (\frac{\bar{a}-u}{\sigma})\sqrt{n_1/(2+m^2w'^2n)}$ and $t_2 = (\frac{\bar{b}-u}{\sigma})\sqrt{n/(2+m^2n)}$

For the standard gain we note that $Z_3 = (Y-u)/\sigma$. has partial regressions on Z_1 and Z_3. Thus $E(Z_3) = \beta_1E(Z_1)+\beta_2E(Z_2)$. The partial regressions β_1 and β_2 can be obtained from the variance covariance matrix given above and using the results of section I.1. equation (8-6) we can write

$$E(Z_1) = \frac{1}{P}\frac{1}{\sqrt{2\pi}}\left[e^{-t_1^2/2}I(t_1) + \rho\ e^{-t_2^2/2}I(t_2)\right] \text{ and}$$

$$E(Z_2) = \frac{1}{P\sqrt{2\pi}}\left[\rho\ e^{-t_1^2/2}I(t_1) + e^{-t_2^2/2}I(t_2)\right]$$

IV.1. Exponential

In this section results are obtained on the assumption that the varietal means follow exponential distribution with parameter λ and the results of the first stage are also taken

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

into consideration. As in (8-20) Probability of a variety with mean Y to be selected in two stages is

$$I_{2,r}\left(\frac{\bar{a}-Y}{\sqrt{\frac{2}{n_1}\sigma}}, \frac{\bar{b}-Y}{\sqrt{\frac{2}{n}\sigma}}\right) \tag{8-23}$$

(8-23) can be written in either of the following two forms

$$\frac{1}{(1-r^2)^{\frac{1}{2}}2\pi} \int_{\frac{\bar{b}-Y}{E_2}}^{\infty} e^{-Z_2^2/2} \left[\int_{\frac{\bar{a}-Y}{E}}^{\infty} e^{-(rZ_2-Z_1)^2/2(1-r^2)}\, dZ_1 \right] dZ_2 \tag{8-24}$$

$$\text{or} \quad \frac{1}{2\pi(1-r^2)^{\frac{1}{2}}} \int_{\frac{\bar{a}-Y}{E}}^{\infty} e^{-Z_1^2/2} \left[\int_{\frac{\bar{b}-Y}{E_2}}^{\infty} e^{-\frac{(rZ_1-Z_2)^2}{2(1-r^2)}}\, dZ_2 \right] dZ_1 \tag{8-25}$$

From that we will get the fraction selected for the exponential distribution.

$$\int_0^{\infty} \lambda\, e^{-\lambda Y} I_{2,r}\left[\left(\frac{\bar{b}-Y}{E_2}\right), \frac{\bar{a}-Y}{E}\right] dY \tag{8-26}$$

Integrating by parts and using the forms (8-24) and (8-25) gives the fraction selected

$$I_{2,r}\left(\frac{\bar{b}}{E_2}, \frac{\bar{a}}{E}\right) + \frac{1}{2\pi(1-r^2)^{\frac{1}{2}}E} \int_0^{\infty} e^{-\lambda Y} \cdot e^{-\frac{(\bar{a}-Y)^2}{2E^2}} \int_{\frac{\bar{b}-Y}{E_2}}^{\infty} e^{-(r(\frac{\bar{a}-Y}{E})-Z_2)^2/2\cdot(1-r^2)} dZ_2 dY$$

$$+ \frac{1}{2\pi(1-r^2)^{\frac{1}{2}}} \frac{1}{E_2} \int_0^{\infty} e^{-\lambda Y} e^{-\frac{(\bar{b}-Y)^2}{2E_2^2}} \int_{\frac{\bar{a}-Y}{E}}^{\infty} e^{-(r(\frac{\bar{b}-Y}{E_2})-Z_1)^2/2(1-r^2)} dZ_1 dY \tag{8-27}$$

First term in (8-27) is a bivariate integral. Second and third terms are also bivariate integrals but they need to be transformed into proper forms. Solution of the second term is indicated below and that of the third can be written down by noting the symmetry. Second term may be written as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{1}{2\pi(1-r^2)^{\frac{1}{2}}EE_2}\int_0^\infty \bar{e}^{\lambda Y}\cdot\bar{e}^{(\bar{a}-Y)^2/2E^2}\int_0^\infty \left\{r\left(\frac{\bar{a}-Y}{E}\right)-\frac{Z+b-Y}{E_2}\right\}^2\cdot\frac{1}{(1-r^2)2}\,dZ\,dY$$

After going through some algebra it can be written

$$\frac{e^{(\lambda^2E^2-2\lambda\bar{a})/2}}{2\pi(1-\rho^2)^{\frac{1}{2}}\sigma_Z\sigma_Y}\int_0^\infty\int_0^\infty \exp-\tfrac{1}{2}\left[\frac{(Z-u_Z)^2}{\sigma_{Z^2}}+\frac{(Y-u_Y)^2}{\sigma_{Y^2}}-2\rho\frac{(Z-u_Z)(Y-u_Y)}{\sigma_Z\quad\sigma_Y}\right]\frac{1}{1-\rho^2}\,dZ\,dY \tag{8-28}$$

Where $u_y=(\bar{a}-E^2\lambda)$ and $u_z=(\bar{a}-\bar{b}-\lambda E^2+EE_2r)$: σ_z and σ_y are given by the covariance matrix

Z	Y
$E^2+E_2{}^2-2EE_2r$	$(E_1-E_2r)E$
-	E^2

Changing to standardized variables (8-28) can be written

$$\int_{t_2}^\infty\int_{t_1}^\infty f_{2,\rho}(Z_1\ Z_2)\,dZ_1dZ_2\cdot\quad \exp(\lambda^2E^2-2\bar{a}\lambda)/2 \tag{8-29}$$

$t_1=-(\bar{a}-\bar{b}-\lambda E_2{}^2+E\lambda E_2r)/\sqrt{E^2+E_2{}^2-2EE_2r}$ and $t_2=-\left(\frac{\bar{a}-E^2\lambda}{E}\right)$

By symmetry the solution of the third term of (8-27) may be given as

$$\int_{t_2'}^\infty\int_{t_1'}^\infty f_{2,\rho'}(Z_1\ Z_2)\,dZ_1\,dZ_2\quad\exp(\lambda^2E_2{}^2-2\lambda\bar{b})/2 \tag{8-30}$$

here $\rho'=\frac{(E_2-Er)}{\sqrt{E^2+E_2{}^2-2EE_2r}}$ $\quad t_1'=\frac{-(\bar{b}-\bar{a}-\lambda E^2+EE_2\lambda r)}{\sqrt{E^2+E_2{}^2-2EE_2r}}$ $\quad t_2'=\frac{-(\bar{b}-E_2{}^2\lambda)}{E_2}$

To obtain the gain we again work in terms of the standard gain

$$=E(\lambda Y-1)=\int_0^\infty Y\cdot\lambda^2\bar{e}^{\lambda Y}\,I_{2,r}\left(\frac{\bar{b}-Y}{E_2},\ \frac{\bar{a}-Y}{E}\right)dY/P\ -1 \tag{8-31}$$

Utilizing the forms (8-24) and (8-25) of $I_{2,r}(..)$ the above expression may be written as

$$\frac{1}{P2\pi\quad E}\int_0^\infty Y\lambda\bar{e}^{\lambda Y}\bar{e}^{\frac{(\bar{a}-Y)^2}{2E^2}}\int_{\frac{\bar{b}-Y}{E}}^\infty \bar{e}^{\frac{1}{2}}\left\{r\left(\frac{\bar{a}-Y}{E}\right)-Z_2\right\}^2/(1-r^2)\,dZ_2\,dY \tag{8-32}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$+\frac{1}{P\ 2\pi\ E_2}\int_0^{\infty} e^{-\lambda Y}\lambda Y\ e^{-(b-Y)^2/2E_2^2}\int_{\frac{\bar{a}-Y}{E}}^{\infty} e^{-\frac{1}{2}\left\{\frac{r(\bar{b}-Y)-Z_1}{E_2}\right\}^2/(1-r^2)}dZ_1dY$$

The two terms of (8-32) are taken up separately. Compare the first term of (8-32) - call it C_1 - with the second term of (8-27) - call that C. Ignoring the constant P, C_1 is the expectation of Y with respect to C and noting the solution of C in (8-29) C_1 can be written as

$$\frac{\lambda\ e^{\frac{1}{2}(\lambda^2E^2-2\bar{a}\lambda)}}{P}\int_{t_2}^{\infty}E_1Z_2\int_{t_1}^{\infty}f_{2,\rho}(--)dZ_1dZ_2+(C)(\bar{a}-\lambda E^2)\frac{\lambda}{P} \qquad (8\text{-}33)$$

t_1, t_2 and ρ have been defined in the solution of C. Similarly comparing the second term of (8-32) - call it D_1 - with the third term of (8-27) - call it D - and noting the solution of D in (8-30) D_1 can be expressed as

$$\frac{e^{\frac{1}{2}(\lambda^2E_2^2-2\lambda\bar{b})}}{P}\int_{t_2'}^{\infty}E_2Z_2\int_{t_1'}^{\infty}f_{2,\rho'}(..)dZ_1dZ_2+D(b-E_2)\frac{\lambda}{P} \qquad (8\text{-}34)$$

Thus in principle the problem is reduced to solve the integral of the form $\int_{t_2}^{\infty}\int_{t_1}^{\infty}Z_2f_{2,\rho}(..)dZ_1dZ_2$ and the solution of that has been given in section II.1.

IV.2. Double exponential

Finally we derive formulae for the double exponential when the results of the 1st stage are considered along with the 2nd stage result,

Prob (V_i, with mean yielding capacity Y)$=I_{2,r}\left(\frac{\bar{b}-Y}{E_2},\frac{\bar{a}-Y}{E}\right)$

That will give the fraction of the population selected in two stages

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$= \tfrac{1}{2} \int_{-\infty}^{\infty} \lambda e^{\lambda|Y-u|} I_{2,r}\left(\frac{\bar{b}-Y}{E_2}, \frac{b-Y}{E}\right) dY$$

$$= \tfrac{1}{2} \int_{u}^{\infty} \lambda e^{\lambda u} e^{-\lambda Y} I_{2,r}\left(\frac{\bar{b}-Y}{E_2}, \frac{\bar{a}-Y}{E}\right) dY + \tfrac{1}{2} \int_{-\infty}^{u} \lambda e^{-\lambda u} e^{\lambda Y} I_{2,r}\left(\frac{\bar{b}-Y}{E_2}, \frac{\bar{a}-Y}{E}\right) dY \qquad (8\text{-}34)$$

$$= \tfrac{1}{2}\left[(A) + (B)\right]$$

$$(A) = I_{2,r}\left(\frac{\bar{b}-u}{E_2}, \frac{\bar{a}-u}{E}\right) + \frac{e^{\lambda u}}{D\ E} \int_{u}^{\infty} e^{\lambda Y} \int_{\frac{b-Y}{E_2}}^{\infty} e^{\frac{1}{2}\left[\frac{\left(\frac{\bar{a}-Y}{E}\right)^2 + Z_2^2 - 2rZ_2\left(\frac{\bar{a}-Y}{E}\right)}{1-r^2}\right]} dZ_2 dY$$

$$\frac{+e^{\lambda u}}{\sqrt{D}\,E_2} \int_{u}^{\infty} e^{\lambda Y} \int_{\frac{\bar{a}-Y}{E}}^{\infty} \exp -\tfrac{1}{2}\left[\frac{\left(\frac{\bar{b}-Y}{E_2}\right)^2 + Z_1^2 - 2rZ_1\left(\frac{\bar{b}-Y}{E_2}\right)}{1-r^2}\right] dZ_1 dY$$

$$(B) = I_{2,r}\left(\frac{\bar{b}-u}{E_2}, \frac{\bar{a}-u}{E}\right) - \frac{e^{\lambda u}}{|D|\ E} \int_{-\infty}^{u} e^{\lambda Y} \int_{-\infty}^{\frac{Y-\bar{b}}{E_2}} \exp -\tfrac{1}{2}\left[\frac{\left(\frac{\bar{a}-Y}{E}\right)^2 + Z_2^2 - 2rZ_2\left(\frac{Y-\bar{a}}{E}\right)}{1-r^2}\right] dZ_2 dY$$

$$- \frac{e^{\lambda u}}{|D|\ E_2} \int_{-\infty}^{u} \int_{-\infty}^{\frac{Y-\bar{a}}{E}} \exp -\tfrac{1}{2}\left[\frac{\left(\frac{Y-\bar{b}}{E_2}\right)^2 + Z_1^2 - 2rZ_1\left(\frac{Y-\bar{b}}{E_2}\right)}{1-r^2}\right] dZ_1\, dY$$

$$|\bar{D}| = \frac{1}{2\pi(1-r^2)^{\frac{1}{2}}}$$

After going through some algebra we can write the fraction selected as,

$$\tfrac{1}{2}\Big\{2I_{2,r}\left(\frac{\bar{b}-u}{E_2}, \frac{\bar{a}-u}{E}\right) + I_{2,\rho}(t_1, t_2)\, e^{\frac{\lambda^2E^2}{2} - \lambda(\bar{a}-u)} + I_{2,\rho'}(t_1', t_2')\, e^{\frac{\lambda^2E_2^2}{2} - \lambda(\bar{b}-u)}$$

$$t_1 = \frac{-(a-b-\lambda E^2 + E\ E_2\lambda r)}{\sqrt{E^2+E_2^2-2EE_2r}}$$

$$t_2 = (u-\bar{a}+E^2\lambda)/E$$

$$\rho = \frac{(E-E_2r)}{\sqrt{E^2+E_2^2-2EE_2r}}$$

$$t_1' = \frac{-(b-a-\lambda E_2^2 + EE_2\lambda r)}{\sqrt{E^2+E_2^2-2EE_2r}}$$

$$t_2' = (u-\bar{b}+E_2^2\lambda)/E_2$$

$$\rho' = \frac{(E_2-Er)}{\sqrt{E^2+E_2^2-EE_2r}}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$+F_{2,R}(\ell_1\ell_2)e^{\frac{\lambda^2E^2}{2}+\lambda(\bar{a}-u)} \qquad \ell_1=\frac{-(\bar{b}-\bar{a}-\lambda E^2+EE_2\lambda r)}{\sqrt{E^2+E_2^2-2rE\,E_2}}$$

$$\ell_2=\frac{u-a-\lambda E^2}{E}$$

$$+F_{2,R'}(\ell_1',\ell_2')e^{\frac{\lambda^2E^2}{2}+\lambda(\bar{b}-u)}\Big\} \qquad R=\frac{-(E-E_2r)}{\sqrt{E^2+E_2^2-2rEE_2}}$$

and finally

$$\ell_1'=\frac{-(a-b-\lambda E_2^2+EE_2\lambda r)}{\sqrt{(E^2+E^2-2rEE_2)}}$$

$$\ell_2'=\frac{u-\bar{b}-\lambda E_2^2}{E_2}$$

$$R'=\frac{-(E_2-E\,r)}{\sqrt{E_1^2+E_2^2-2EE_2r}} \qquad (8\text{-}35)$$

For the gain we again work in terms of standard gain $E(Y-u)/(\sqrt{2}/\lambda)$.

$$=\frac{1}{P\;2\sqrt{2}}\left\{\int_u^\infty \lambda^2Y\; e^{\lambda u}\; e^{-\lambda Y} I_{2,r}\left(\frac{\bar{b}-Y}{E_2},\frac{\bar{a}-Y}{E}\right)dY\right.$$

$$\left.+\int_{-\infty}^{u}\lambda^2Y\; e^{\lambda Y}\; e^{-\lambda u}\; I_{2,r}\left(\frac{\bar{b}-Y}{E_2},\frac{\bar{a}-Y}{E}\right)dY\right\}-\frac{\lambda u}{\sqrt{2}}$$

$$=\frac{1}{2\sqrt{2}\;P}\left[\;A+B\;\right]-\frac{\lambda u}{\sqrt{2}}$$

$$(A)=\lambda u I_{2,r}\left(\frac{\bar{b}-u}{E_2},\frac{\bar{a}-u}{E}\right)+\lambda\; e^{\lambda u}\int_u^\infty e^{-\lambda Y} I_{2,r}\left(\frac{\bar{a}-Y}{E},\frac{b-Y}{E_2}\right)dY$$

$$+\frac{\lambda e^{\lambda u}}{|D|\,E}\int_u^\infty \bar{e}^{\lambda Y}\int_{\frac{b-Y}{E_2}}^\infty \exp-\tfrac{1}{2}\left|\left(\frac{\bar{a}-Y}{E^2}\right)^2+Z_2^2-2r(Z_2)\left(\frac{\bar{a}-Y}{E}\right)\right]\frac{1}{1-r^2}\;dZ_2dY$$

$$+\frac{\lambda\; e^{\lambda u}}{|D|\cdot E_2}\int_u^\infty \bar{e}^{\lambda Y}\int_{\frac{\bar{a}-Y}{E}}^\infty \exp-\tfrac{1}{2}\left|\left(\frac{\bar{a}-Y}{E^2}\right)^2+Z_1^2-2r\;Z_1\left(\frac{\bar{a}-Y}{E}\right)\right]\frac{1}{1-r^2}\;dZ_1dY$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$(B)=\lambda u\, I_{2,r}\left(\frac{\bar{b}-u}{E_2},\frac{\bar{a}-u}{E}\right)-\lambda\,\bar{e}^{\lambda u}\int_{-\infty}^{u} e^{\lambda Y} I_{2,r}\left(\frac{\bar{a}-Y}{E},\frac{\bar{b}-Y}{E_2}\right)$$

$$-\frac{\lambda\,\bar{e}^{\lambda u}}{E}\int_{-\infty}^{u}\int_{-\infty}^{\frac{Y-\bar{b}}{E_2}} Y\, e^{\lambda Y}\exp-\tfrac{1}{2}\left[\frac{(Y-\bar{a})^2}{E^2}+Z_2^2-2rZ_2\left(\frac{Y-\bar{a}}{E}\right)\right]\frac{1}{1-r^2}\,dZ_2\,dy$$

$$-\frac{\lambda\,\bar{e}^{\lambda u}}{E_2}\int_{-\infty}^{u}\int_{-\infty}^{\frac{Y-\bar{a}}{E}} Y\, e^{\lambda Y}\exp-\tfrac{1}{2}\left[\frac{(Y-\bar{b})^2}{E_2^2}+Z_1^2-2rZ_1\left(\frac{Y-\bar{b}}{E_2}\right)\right]\frac{1}{1-r^2}\,dZ_2\,dy$$

A little algebra shows that

$$\text{Standard gain}=\frac{1}{2\cdot\sqrt{2}\cdot P}\Big[2\lambda u\, I_{2,r}\left(\frac{\bar{b}-u}{E_2},\frac{\bar{a}-u}{E}\right)+e^{\frac{\lambda^2E^2}{2}-\lambda(\bar{a}-u)}I_{2,\rho}(t_1,t_2)$$

$$+e^{\frac{\lambda^2E^2}{2}-\lambda(\bar{b}-u)}I_{2,\rho'}(t_1',t_2')+e^{\frac{\lambda^2E^2}{2}-\lambda(\bar{a}-u)}\left\{\lambda E I_{2,\rho}(t\ t_2-Y)+\lambda(\bar{a}-\lambda E^2)I_{2,\rho}(t_1,t_2)\right\}$$

$$+e^{\frac{\lambda^2E_2^2}{2}-\lambda(\bar{b}-u)}\left\{\lambda E_2 I_{2,\rho'}(t_1',t_2'-Y)+\lambda(\bar{b}-\lambda E_2^2)I_{2,\rho'}(t_1',t_2')\right\}$$

$$+e^{\frac{\lambda^2E^2}{2}+\lambda(\bar{a}-u)}F_{2,R}(\ell_1,\ell_2)+e^{\frac{\lambda^2E_2^2}{2}+\lambda(\bar{b}-u)}F_{2,R'}(\ell_1',\ell_2')$$

$$-e^{\frac{\lambda^2E^2}{2}+\lambda(\bar{a}-u)}\left\{\lambda E F_{2,R}(\ell_1,\ell_2-Y)+\lambda(\bar{a}+\lambda E^2)F_{2,R}(\ell_1,\ell_2)\right\}\qquad(8\text{-}36)$$

$$-e^{\frac{\lambda^2E_2^2}{2}+\lambda(\bar{b}-u)}\left\{\lambda E_2 F_{2,R'}(\ell_1',\ell_2'-Y)+\lambda(\bar{b}+\lambda E_2^2)F_{2,R'}(\ell_1',\ell_2')\right\}\Big]-\frac{\lambda u}{\sqrt{2}}$$

The constants t_1, t_2, ρ, etc. have been defined in the solution of P, the fraction selected.

It may be noted that from the formulae for the second stage when the results of 1st stage are considered we can get the corresponding formulae when the results of the 1st stage are not considered by taking 'r' equal to zero.

Having obtained the formulae for the two stage screening we get the numerical results. No numerical results are obtained for the double exponential; the formulae derived for the double exponential are not very useful for numerical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

calculations, because they will introduce too much error due to rounding off. In a two stage selection the results obtained will depend upon n_1 and n_2, the replications at the first and the second stage respectively and 'a' and 'b'. These four quantities can be taken in an infinite number of combinations. Here we take 'a' and 'b' as equal and n_2, the replications at the second stage are increased in such a way that the same area (resources) is used at the second stage as at the first stage. This would amount to say that $n_2 = \frac{n_1}{P}$, where P is the fraction selected at the 1st stage.

As expected the fraction selected is less when the results of the 1st stage are not considered but, the standard gain is more. At lower values of $\frac{\bar{a}-u}{E}$, the fraction selected is higher for the normal but, at higher values, the fraction selected is higher for the exponential. The standard gain is higher for the exponential for the range studied.

Two Stage Selection

Normal distribution

Table 16

Fraction Selected

	Without considering the 1st stage result		Considering the 1st stage result	
$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	.5	1.0	.5	1.0
.1	.25248	.3189	.3240	.3681
.2	.2153	.2908	.2797	.3374
.3	.1805	.2636	.2370	.3073
.5	.1194	.2127	.1598	.2489
.75	.0618	.1566	.0842	.1727
1.0	.0261	.1096	.0356	.1285

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 17

Standard gain

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E}$	Without considering the 1st stage		Considering the 1st stage	
	.5	1.0	.5	1.0
.1	.8604	1.1518	.5436	.7019
.2	.9415	1.2253	.60840	.7444
.3	1.0194	1.2959	.6809	.7932
.5	1.2357	1.4443	.8523	.9123
.75	1.5461	1.6011	1.1828	1.0652
1.0	1.933	1.8613	1.4826	1.2672

Table 18

Two stage Selection

Exponential Distribution - Fraction Selected

$\frac{\bar{a}-u}{E}$ \ $\frac{\sigma_o}{E} = \frac{1}{\lambda E}$	Without considering the 1st stage		Considering the 1st stage	
	.5	1.0	.5	1.0
.1	.2304	.2576	.2815	.3127
.2	.1963	.2350	.2479	.2947
.3	.1646	.2098	.2147	.2704
.5	.1131	.1728	.1766	.2462
.75	0.0660	.1342	.1411	.1783
1.0	.0384	.1060	.1128	.1370
1.5	.0130	.0630	.0619	.0829

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 19

Standard gain

$\frac{a-u}{E}$ \ $\frac{\sigma_o}{E} = \frac{1}{\lambda E}$	Without considering the 1st stage		Considering the 1st stage	
	.5	1.0	.5	1.0
.1	.8916	1.1518	.6165	.9000
.2	.9511	1.2532	.7529	.9820
.3	1.2094	1.3674	.8277	1.050
.5	1.580	1.570	.9917	1.1871
.75	2.1609	1.837	1.3024	1.4851
1.0	2.775	2.007	1.654	1.857
1.5	4.0166	2.8199	1.8404	2.2743

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 9

Screening on Two Characters

Introduction

I. In several cases it may be desirable to base screening on more than one character. For example in plant selection the screening may be based on yield, drought resistance, disease resistance etc. For wheat varieties the bread making quality is as important as the yield. In drug screening the selection may be based upon the primary effect of the drug but at the same time the secondary effects it produces on the subject should be taken into consideration. The whole field of screening on more than one character is unexplored. One possible approach is that one character is of prime importance but other characters are supposed to be above (below) certain specified levels. For example, in wheat only those varieties are selected which are above a certain level in bread making quality, in drug screening only those effective drugs are selected which have side effects less than a certain level. The number of characters considered can be many but, the evaluation of the consequences may become pretty impossible. In the following sections a procedure for selecting on two characters is considered. Only those individuals are passed to the next stage which are above a certain level in one character and in the second character they are better than the standard (control) by a certain specified amount.

First Stage

II. In the parent population the characters may be denoted by Y_1 and Y_2. Y_1 and Y_2 individually follow normal distributions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with variances $\sigma_{1_o}^2$ and $\sigma_{2_o}^2$, means u_1 and u_2 respectively. They may or may not be correlated.

$X_{1ij} = u_{1i} + e_{1ij}$ - observation for character Y_1 on Variety V_i

$X_{2ij} = u_{2i} + e_{2ij}$ - observation for character Y_2 on Variety V_i

$X_{2oj} = u_{2o} + e_{2ij}$ - observation for character Y_2 on the Control.

e_{1ij} and e_{2ij} are normally distributed with variances σ_1^2 and σ_2^2 respectively. Variety V_i will be selected only if

$$\bar{X}_{1i} = \sum_{j=1}^{n_1} \frac{X_{1ij}}{n_1} > q \qquad (q \text{ a positive constant})$$

and $$\bar{X}_{2i} - \bar{X}_{2o} = \tilde{X}_{2i} = \sum_{j=1}^{n_1} \frac{X_{2ij}}{n_1} - \sum_{j=1}^{n_1} \frac{X_{2oj}}{n_1} > a\delta$$

Four different cases may be distinguished

(1) Y_1 and Y_2 are not related and e_{1ij} and e_{2ij} are also not related.

(2) Y_1 and Y_2 are independent but e_{1ij} and e_{2ij} are correlated with correlation r'.

(3) Y_1 and Y_2 are related with correlation ρ and e_{1ij} and e_{2ij} are independent.

(4) Y_1 and Y_2 have a correlation ρ and e's are related with correlation r'.

Variance of $X_{1ij} = \mathrm{Var}(u_{1i} + e_{1ij}) = \sigma_1^2$

Variance of $X_{2ij} = \mathrm{Var}(u_{2i} + e_{2ij}) = \sigma_2^2$

The first three types are special cases of the last one. From the general case any of the others can be obtained by equating ρ, r or both to zero.

$$\text{correlation } (\bar{X}_{1i}, \tilde{X}_{2i}) = \frac{r'}{\sqrt{2}} = r$$

$$\text{Variance } (\bar{X}_{1i}) = \frac{\sigma_1^2}{n_1} = E_1^2 \text{ and } \mathrm{Var}(\tilde{X}_{2i}) = \frac{2\sigma_2^2}{n_1} = E_2^2$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

With this procedure the probability of a variety with means Y_1 and Y_2 for the first and second character respectively to be selected will be

$$\frac{1}{2\pi\,(1-r^2)^{\frac{1}{2}}E_1E_2}\int_{a\delta}^{\infty}\int_{q}^{\infty}\exp-\tfrac{1}{2}\left\{\frac{(X_1-Y_1)^2}{E_1^{\,2}}-\frac{(X_2-Y_2+u_0)^2}{E_2^{\,2}}-\frac{2r(X_1-Y_1)(X_2-Y_2+u_0)}{E_1E_2}\right\}\frac{1}{1-r^2}\,dX_1\,dX_2$$

(9-1)

Y_1 and Y_2 are supposed to follow a bivariate distribution with variances $\sigma_{1_0}^{\,2}$ and $\sigma_{2_0}^{\,2}$; means u_1 and u_2 and correlation ρ. That will give the fraction of the population selected

$$\frac{1}{\sigma_{1_0}\sigma_{2_0}(2\pi)^2E_1E_2(1-r^2)^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{a\delta}^{\infty}\int_{q}^{\infty}\exp-\tfrac{1}{2}\left[\left\{\frac{(Y_1-u_1)^2}{\sigma_{1_0}^{\,2}}+\frac{(Y_2-u_2)^2}{\sigma_{2_0}^{\,2}}-\frac{2\rho(Y_1-u_1)(Y_2-u_2)}{\sigma_{1_0}\sigma_{2_0}}\right\}\frac{1}{1-\rho}\right.$$

$$\left.+\frac{1}{1-r^2}\left\{\frac{(X_1-Y_1)^2}{E_1^{\,2}}+\frac{(X_2-Y_2+u_0)^2}{E_2^{\,2}}-\frac{2r(X_1-Y_1)(X_2-Y_2+u_0)}{E_1E_2}\right\}\right]dX_1\,dX_2\,dY_1\,dY_2$$

(9-2)

In (9-2) substitute

$$\sqrt{\frac{n_1}{2}}\,\frac{(X_2-u_2+u_0)}{\sigma_2}=Z_2,\qquad \sqrt{n_1}\,\frac{(X_1-u_1)}{\sigma_1}=Z_1$$

$$\sqrt{\frac{n_1}{2}}\,\frac{(Y_2-u_2)}{\sigma_{2_0}}=t_2\qquad \sqrt{n_1}\,\frac{(Y_1-u_1)}{\sigma_{1_0}}=t_1,\ \text{then (9-2)}$$

reduces to

$$\frac{\sqrt{2}}{(2\pi)^2n_1(1-r^2)^{\frac{1}{2}}(1-\rho^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{\frac{(a-u_2)}{E_2}}^{\infty}\int_{\frac{(b-u_1)}{E_1}}^{\infty}\exp-\tfrac{1}{2}\left(\frac{t_1^2}{n_1}+\frac{t_2^2}{n_1}-\frac{2\rho t_1t_2\sqrt{2}}{n_1}\right)\frac{1}{1-\rho^2}$$

$$+\left\{\frac{(\sigma_2Z_2-t_2\sigma_{2_0})^2}{\sigma_2^{\,2}}+\frac{(\sigma_1Z_1-\sigma_{1_0}t_1)^2}{\sigma_1^{\,2}}-\frac{2r(\sigma_1Z_1-\sigma_{1_0}t_1)(\sigma_2Z_2-\sigma_{2_0}t_2)}{\sigma_1\sigma_2}\right\}\frac{1}{1-r^2}\Bigg]dZ_1\,dZ_2\,dt_1\,dt_2$$

(9-3)

(9-3) is a four variate normal integral with variance covariance matrix.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Z_1	Z_2	t_1	t_2
$1 + \frac{\sigma_{1_o}^2}{E_1^2}$	$r - \frac{\sigma_{2_o}\sigma_{1_o}}{\sigma_2\sigma_1}\frac{\rho\, n_1}{\sqrt{2}}$	$\frac{n\sigma_{1_o}}{\sigma_1}$	$-\rho\frac{\sigma_{1_o}}{\sigma_1}\frac{n_1}{\sqrt{2}}$
	$1 + \frac{\sigma_{2_o}^2}{E_2^2}$	$\frac{-n_1\sigma_{2_o}\rho}{\sigma_2\sqrt{2}}$	$\frac{\sigma_{2_o}}{\sigma_2}\frac{n_1}{2}$
		n_1	$-\frac{n_1\rho}{\sqrt{2}}$
			$\frac{n_1}{2}$

As noted before the variances for any of the first three cases will be given by equating r, ρ or both to zero in the above matrix. Integrate t_1 and t_2 and we are left with a bivariate integral variances and covariance for which will be given by the left hand top corner submatrix shown above. And changing to standardized variables it will take the form

$$\frac{1}{2\pi\,(1-R^2)^{\frac{1}{2}}}\int_{t_2}^{\infty}\int_{t_1}^{\infty}\exp -\tfrac{1}{2}(Z_1^2+Z_2^2-2RZ_1\,Z_2)/(1-R^2)\;dZ_1dZ_2 = P, \text{ say}$$

$t_2=(\bar{a}-u_2)/\sqrt{(E_2^2+\sigma_{2_o}^2)}$, $t_1=(q-u_1)/\sqrt{(E_1^2+\sigma_{1_o})}$; R is the correlation determined from the above submatrix.

To find the gain either in Y_1 or Y_2 we can proceed as follows. Method for Y_2 is indicated. For Y_1 it will be just the same with obvious modifications in substitutions. Changing to standardized variables (9-3) may be written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty} f(x_1,x_2,x_3,x_4)\;dx_1dx_2dx_3dx_4 \qquad (9\text{-}4)$$

$x_4 = \frac{Y_2-u_2}{\sigma_{2_o}}$, thus $E(x_4)$ is equal to the gain in Y_2 in units of standard deviation or the standard gain in Y_2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$= \frac{1}{P} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{t_2}^{\infty} \int_{t_1}^{\infty} x_4 \, f(x_1,x_2,x_3,x_4) \, dx_1 dx_2 dx_3 dx_4 \qquad (9\text{-}5)$$

$$= \frac{1}{P} \int_{-\infty}^{\infty} \int_{t_2}^{\infty} \int_{t_1}^{\infty} x_4 f(x_1,x_2,x_4) \, dx_1 \, dx_2 \, dx_4 \qquad (9\text{-}6)$$

x_4 has regression on x_1 and x_2. The partial regressions of x_4 on x_1 and x_2 can be obtained from the covariance matrix. By virtue of the regressions $E(x_4)=\beta_2 E(x_2) + \beta_1 E(x_1)$. The procedure to find $E(x_1)$ and $E(x_2)$ has been indicated in section I of chapter 8.

III. Second Stage

At the second stage again two alternatives are possible. Either the results of the first stage may be taken into consideration or the selection may be based upon the results of the second stage only. Both the cases can be dealt with simultaneously by taking into consideration the following explanation of the terms involved.

At the second stage the selection will be based upon $n_1+n_2=n$ replications, n_1 is the number of replications at the first stage and n_2 those at the second stage and obviously n is the sum of the two. If at the second stage the result of the second stage only is considered the same notation may be used with the understanding that $n=n_2$ only. Secondly, in the absence of the result of the first stage being taken into consideration the factor $\sqrt{n_1/n}$ (to be explained below) should be equated to zero. These two alternatives may be considered in conjunction with any of the four types listed in the first stage selection.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

With the above explanation in mind it will be sufficient to indicate the procedure for the case when the result of the first stage is considered and that is associated with type (4) of the first stage. Other types are only special cases of this general formulation.

At the second stage variety V_i will be selected if

$$\bar{\bar{X}}_{1i} = \sum_{j=1}^{n} \frac{X_{1ij}}{n} > q$$

$$\tilde{\tilde{X}}_{2i} = \bar{\bar{X}}_{2i} - \bar{\bar{X}}_{2o} = \sum_{j=1}^{n} \frac{(X_{2ij} - X_{2oj})}{n} > b\delta$$

Variance $(\bar{\bar{X}}_{1i}) = \sigma_1^2/n$ and $\mathrm{Var}(\tilde{\tilde{X}}_{2i}) = 2\sigma_2^2/n$. Thus the factors incluencing the selection or rejection of V_i are $\bar{X}_{1i}$, $\bar{\bar{X}}_{1i}$, $\tilde{X}_{2i}$ and $\tilde{\tilde{X}}_{2i}$. These four quantities are correlated with the correlatinn structure

$\bar{X}_{1i}=X_1$	$\bar{\bar{X}}_{1i}=X_2$	$\tilde{X}_{2i}=t_1$	$\tilde{\tilde{X}}_{2i}=t_2$	
1	c	r	rc	
	1	rc	r	
		1	c	$c = \sqrt{n_1/n}$, r is the same as defined in 1st stage
			1	

Thus the probability that a variety with means Y_1 and Y_2 for the two characters respectively will be selected at the second stage will be given by a four variate normal integral with the correlation structure given in the above matrix, and can be written as

$$\int_{\frac{\bar{b}-Y_2}{\sigma_2\sqrt{2/n}}}^{\infty} \int_{\frac{\bar{a}-Y_2}{\sigma_2\sqrt{2/n}}}^{\infty} \int_{\frac{q-Y_1}{\sigma_1/\sqrt{n}}}^{\infty} \int_{\frac{q-Y_1}{\sigma_1/\sqrt{n}}}^{\infty} f(X_1, X_2, t_1, t_2)\, dX_1 dX_2 dt_1 dt_2 = I(\ldots) \qquad (9\text{-}7)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The above expression obviously depends upon Y_1 and Y_2. In the parent population Y_1 and Y_2 are supposed to follow a bivariate normal distribution. Thus the fraction of the population selected at the second stage will be

$$\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{\exp -\frac{1}{2(1-\rho^2)}(Y_1^2 + Y_2^2 - 2\rho Y_1 Y_2)\right\} I(....)dY_1dY_2 \qquad (9\text{-}8)$$

$I(....)$ refers to the expression in (9-7).

In (9-8) substitute $\sqrt{n}\left(\frac{Y_1-u_1}{\sigma_{1_o}}\right) = W_3$; $\sqrt{n_1}\left(\frac{X_1-u_1}{\sigma_1}\right) = W_1$; $\sqrt{n}\left(\frac{X_2-u_2}{\sigma_1}\right)=W_2$

$$\sqrt{n}\left(\frac{Y_2-u_2}{\sigma_{2_o}}\right)=Z_3; \quad \sqrt{n_1}\left(\frac{t_1-u_2+u_o}{\sigma_2}\right)=Z_1 \quad \sqrt{n}\left(\frac{t_2-u_2+u_o}{\sigma_2}\right)=Z_2$$

With these substitutions (9-8) reduces to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty} f(.....)dW_1dW_2dZ_1dZ_2dW_3dZ_3 \qquad (9\text{-}9)$$

f is a six variate normal density function with variance covariance matrix

W_1	W_2	Z_1	Z_2	W_3	Z_3
$\frac{1+n\sigma_{1_o}^2c^2}{\sigma_1^2}$	$c+\frac{n\sigma_{1_o}^2 \cdot c}{\sigma_1^2}$	illegible	illegible	$cn\frac{\sigma_{1_o}}{\sigma_1}$	$\frac{\rho}{\sqrt{2}}\, c\, \frac{\sigma_{1_o} n}{\sigma_1}$
	$1+\frac{n\sigma_{1_o}^2}{\sigma_1^2}$	illegible	illegible	$\frac{\sigma_{1_o}n}{\sigma_1}$	$\frac{\rho}{\sqrt{2}}\ \frac{\sigma_{1_o}n}{\sigma_1}$
		illegible	illegible	$\frac{c\rho}{\sqrt{2}}\, n\, \frac{\sigma_{2_o}}{\sigma_2}$	$c\ \frac{\sigma_{2_o}n}{\sigma_2 2}$
			$1+\frac{\sigma_{2_o}^2 n}{\sigma_2^2 2}$	$\frac{\rho\sigma_{2_o}n}{\sqrt{2}\ \sigma_2}$	$\frac{\sigma_{2_o}}{\sigma_2}\ \frac{n}{2}$
				n	$\frac{\rho\, n}{\sqrt{2}}$
					$\frac{n}{2}$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Again transforming W's and Z's to standardized variables (denoting the standardized variables by the same letters) we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{t_2'}^{\infty}\int_{t_1'}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty} f(....)\, dW_1 dW_2 dZ_1 dZ_2 dW_3 dZ_3$$

$$t_1 = \frac{(q-u_1)\sqrt{n}}{\sqrt{\sigma_1^2 + n\sigma_{1_o}^2 c^2}} \qquad t_2 = \frac{(q-u_1)\sqrt{n}}{\sqrt{\sigma_1^2 + n\sigma_{1_o}^2}} \qquad t_1' = (\bar{a}-u_2)\sqrt{n_1}/\sqrt{2\sigma_2^2 + c^2\sigma_{2_o}^2 n}$$

$$t_2' = (\bar{b}-u_2)\sqrt{n}/\sqrt{2\sigma_2^2 + n\sigma_{2_o}^2}$$

Integrating out W_3 and Z_3 we get the fraction selected

$$I(t_1, t_2, t_1, t_2) = P, \text{ say}$$

This is a four variate normal integral with correlation structure given by the left hand top corner submatrix. $Z_3 = (Y_2 - u_2)/\sigma_{2_o}$, hence $E(Z_3)$ is equal to the standard gain in Y_2 after two stage screening. Standard gain for Y_2 will be given by the following expression.

$$\frac{1}{P}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{t_2'}^{\infty}\int_{t_1'}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty} Z_3\, f(....)\, dW_1 dW_2 dZ_1 dZ_2 dW_3 dZ_3 \qquad (9\text{-}10)$$

W_3 can be integrated out and we are left with

$$\frac{1}{P}\int_{-\infty}^{\infty}\int_{t_2'}^{\infty}\int_{t_1'}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty} Z_3\, f(....)\, dW_1 dW_2 dZ_1 dZ_2 dZ_3 \qquad (9\text{-}11)$$

To determine the value of the above integral we observe that Z_3 has partial regressions on W_1, W_2, Z_1 and Z_2. These regressions can be obtained from the covariance matrix and $E(Z_3)$ may be written as

$$E(Z_3) = \beta_1 E(Z_1) + \beta_2 E(Z_2) + \beta_3 E(W_1) + \beta_4 E(W_2)$$

At this stage to work with the original variables is pretty tedious. Let us suppose that the inverse of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

correlation matrix for W_1, W_2, Z_1 and Z_2 (they are standardized variables) is of the form

W_1	W_2	Z_1	Z_2
1	a	b	c
	1	d	e
		1	f
			1

If the numerical values are given for the original matrix it is straightforward to calculate the inverse of the left hand corner submatrix. The density function of the four variables will be of the form -

$$\frac{1}{(2\pi)^2\sqrt{|A|}} \exp -\tfrac{1}{2}(W_1^2+W_2^2+Z_1^2+Z_2^2+2aW_1W_2+2bW_1Z_1+2cW_1Z_2+2dW_2Z_1+2eW_2Z_2+2fZ_1Z_2)$$

and that will give

$$E(W_1+aW_2+bZ_1+cZ_2)=\frac{1}{\sqrt{|A|}P(2\pi)^2}\int_{t_2}^{\infty}\int_{t_1}^{\infty}\int_{t_2}^{\infty}\int_{t_1}^{\infty}(W_1+aW_2+bZ_1+cZ_2)\exp -\tfrac{1}{2}(\ldots)\,dW_1\,dW_2\,dZ_1\,dZ_2$$

$$=\frac{1}{\sqrt{|A|}P(2\pi)^2}\int_{t_2}^{\infty}\int_{t_1}^{\infty}\int_{t_2}^{\infty}\exp -\tfrac{1}{2}(t_1^2+W_2^2+\ldots)\,dW_2\,dZ_1\,dZ_2$$

In the exponential the variable W_1 has been replaced by t_1. The resulting integral is a trivariate and with the help of existing tables it can be evaluated. Let us denote it by

$$\frac{1}{P\sqrt{2\pi}}\ I(t_1.t_2t_1't_2') \qquad (9\text{-}12)$$

Similarly $E(W_2+aW_1+dZ_1+eZ_2) = \frac{1}{P\sqrt{2\pi}}\ I(t_2.t_1t_1't_2')$

$$E(Z_1+bW_1+dW_2+fZ_2) = \frac{1}{P\sqrt{2\pi}}\ I(t_1'.t_1t_2t_2')$$

and $$E(Z_2+eW_2+fZ_1+cW_1) = \frac{1}{P\sqrt{2\pi}}\ I(t_2'.t_1t_2t_1')$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From these four relations $E(Z_2)$, $E(Z_1)$, $E(W_1)$ and $E(W_2)$ can be known. Apart from P which is the value of a four variate normal integral all other quantities involved can be evaluated with the help of the existing tables.

Exponential

IV. If the two characters Y_1 and Y_2 follow independently exponential distributions with parameters λ_1 and λ_2 similar expressions as for the normal can be worked out for the two stage selection based upon two characters. They are briefly indicated below.

(a) Following similar steps as in section II the fraction of the population selected at the first stage will be

$$\int_{o}^{\infty}\int_{o}^{\infty} \lambda_1\lambda_2\, e^{-\lambda_1 Y_1 - Y_2\lambda_2} I_{2,r}\left(\frac{\bar{a}-Y_2}{E_2}, \frac{\bar{q}-Y_1}{E_1}\right) dY_1 dY_2 = P \qquad (9\text{-}13)$$

and the standard gain in Y_2 will be given by

$$\frac{1}{P}\int_{o}^{\infty}\int_{o}^{\infty} (\lambda_2 Y_2 - 1)\lambda_1\lambda_2 e^{-\lambda_1 Y_1 - \lambda_2 Y_2} I_{2,r}\left(\frac{\bar{a}-Y_2}{E_2}, \frac{\bar{q}-Y_1}{E_1}\right) dY_1\, dY_2 \qquad (9\text{-}14)$$

(b) Likewise following the steps as in section III for the normal the fraction selected at the second stage will be

$$\int_{o}^{\infty}\int_{o}^{\infty} \lambda_1\lambda_2\, e^{-\lambda_1 Y_1 - \lambda_2 Y_2} I_{4,\text{matx}}\left(\frac{\bar{b}-Y_2}{\sigma_2\sqrt{2/n}}, \frac{\bar{a}-Y_2}{\sqrt{2/n_1}\sigma_2}, \frac{q-Y_1}{\sigma_1\sqrt{1/n}}, \frac{q-Y_1}{\sigma_1\sqrt{1/n_1}}\right) dY_1 dY_2 \qquad (9\text{-}15)$$

and the standard gain after second stage selection is

$$\int_{o}^{\infty}\int_{o}^{\infty} \lambda_2 Y_2 - 1\ \lambda_1 Y_2\ e^{-\lambda Y_1 - \lambda_2 Y_2} I_{4,\text{matx}}\left(\frac{\bar{b}-Y_2}{\sigma_2\sqrt{2/n}}, \frac{\bar{a}-Y_2}{\sqrt{2/n_1}\sigma_2}, \frac{q-Y_1}{\sigma_1\sqrt{1/n}}, \frac{q-Y_1}{\sigma_1\sqrt{1/n_1}}\right) dY_1 dY_2 / \qquad (9\text{-}16)$$

matx implies the correlation structure for the four variables.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Glossary of Symbols

Each symbol is defined when it is first introduced; they are grouped here for easy reference.

$$F(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

$$I(x) = \frac{1}{\sqrt{(2\pi)}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt$$

$$f(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{x^2}{2}}$$

$f_{k\,(\ )}(x_1,..x_k)$ is multivariate density function with correlation structure given in the bracket.

$F_{k\,(\ )}(x_1,..x_k)$ is cumulative distribution function of k variate normal distribution with the correlation structure indicated in the outside bracket. It may be taken as generalization of F(x).

$I_{k,(..)}(x_1,..x_k)$ is the generalization of I(x).

$$F_{2,\rho}(t_1,t_2\text{-}y) = \int_{-\infty}^{t_2}\int_{-\infty}^{t_1} yf_{2,\rho}(x,y)dx\,dy$$

$$I_{2,\rho}(t_1,t_2\text{-}y) = \int_{t_2}^{\infty}\int_{t_1}^{\infty} yf_{2,\rho}(x,y)dx\,dy$$

$f(y/u)$ or $f(y|u)$ is the conditional density function of y for given u.

$$\lambda(t) = \frac{\frac{1}{\sqrt{(2\pi)}} e^{-\frac{t^2}{2}}}{I(t)}$$

E has been used for expectation, such as E(x) and for $\sqrt{(2/n)}\sigma$ also.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bibliography

(1) Anderson, T.W. (1951) "Classification by multivariate analysis", Psychometrika 16; 31-50.

(2) Anscombe, F.J. (1960) "Rejection of outliers", Technometrics 2, 123-147.

(3) Armitage, P. and Maurice Schneiderman (1958) "Statistical problems in mass screening programme". New York Academy of Science - Annals 76; 896-906.

(4) Bahadur, R.R. (1950) "On a problem in the theory of k populations". Annals Mathematical Statistics 21; 362-375.

(5) Bechhoffer, R.E. (1954) "A single sample multiple procedure for ranking means of normal populations with known variance". Annals Mathematical Statistics 25; 16-39.

(6) Bechhoffer, R.E. (1958) "A sequential multiple decision procedure for selecting the best one of several normal populations with a common unknown variance and its use with various experimental designs". Biometrics 14; 408-429.

(7) Bechhoffer, R.E., Dunnett, C.W. and Sobel Milton (1954) "A two sample multiple decision procedure for ranking means of normal populations with common unknown variance". Biometrika 41; 170-176.

(8) Bross Irwin (1950) "Two choice selection". Journal American Statistical Association 45; 530-539.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(9) Birnbaum, Z.W. (1950) "Effect of linear truncation on a multinormal population". Annals Mathematical Statistics 21; 272-279.

(10) Birnbaum, Z.W. and Chapman, D.G. (1950) "On optimum selection from multinormal population". Annals Mathematical Statistics 21; 443-447.

(11) Curnow, R.N. (1958) "The consequences of errors of measurements for selection from certain non-normal distributions". Proceedings of the 31st session of the International Statistical Institute.

(12) Dunnett, C.W. (1955) "A multiple comparison procedure for comparing several treatments with a control". Journal of the American Statistical Association, Vol.50^{2}; 1096-1121.

(13) Cochran, W.G. (1941) "The distribution of the largest of a set of estimated variances as a fraction of the total". Annals Eugenics 11, 47-52.

(14).....(1950) "Improvement by selection". Second Berkley Symposium on Statistics and probability; 449-470.

(15) Davies, O.L. (1958) "The design of screening tests in the pharmaceutical Industry". Bulletin International Institute de Statistique, 36^{3}; 442-448.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(16) Dunnett, C.W. and Milton Sobel (1954) "A bivariate generalization of Student's 't' distribution". Biometrika 41; 153-169.

(17) Dass, S.C. (1956) "The numerical evaluation of a class of integrals - II". Proceedings Cambridge Philosophical Society 52; 442-448.

(18) Curnow, R.N. and Dunnett, C.W. (1962) "Numerical evaluation of certain multivariate integrals". Annals Math. Statistics 33;571-579.

(19) Dunnett, C.W. (1960) "On selecting the largest of k normal populations". Journal of the Royal Statistical Society series B 22; 1-40.

(20) Cadwell, J.H. (1951) "The bivariate normal Integral". Biometrika, 38; 475-481.

(21) Feller, W. (1957) "An introduction to probability theory and its applications". 2nd Ed. New York; Wiley.

(22) Finney, D.J. (1958) "Statistical problems in plant screening". Bull. Institute International de Statistique" 36^3; 242-268.

(23) Fisher 'Sir' R.A. (1959)"Statistical methods and Scientific Inference". London; Oliver and Boyd.

(24) Fisher, R.A. (1936) "The use of multiple measurements in taxonomic problems". Annals Eugenics 7; 179-188.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(25) Finney, D.J. (1956) "The consequences of selection for a variate subject to errors of measurement". Revue Institute International de Statistique 24; 1-9.

(26) Fairfield Smith (1936) "A discriminant function for plant selection". Annals of Eugenics Vol.7; 240-250.

(27) Watson, G.S. (1961) "A study of group screening method". Technometrics Vol.3; 371-388.

(28) Gutman, Irwin (196) "Best populations and tolerance regions". Annals of the Institute of Statistical Mathematics. Vol.XIII, No.1; 9-26.

(29) Grundy, P.M.; Healy, M.J.R. and D.H. Hees (1956) "Economic choice of the amount of experimentation". Journal Royal Statistical Society B18; 32-48.

(30).....(1954) "Decision between two alternatives, How many experiments". Biometrics 10,717-323.

(31) Kendall, M.G. and Stuart, A. (1958) "The advanced theory of Statistics Vol.1" London, Charles Griffin.

(32) Lehman, E.L. (1961) "Some model I problems of selection". Annals Mathematical Statistics, 32; 990-1012.

(33) Mosteller, F. (1948) "A k slippage test for an extreme population". Annals Mathematical Statistics, 19; 58-65.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(34) Neyman, J. (1952) "Lectures and conferences on Mathematical Statistics and Probability". U.S. Department of Agriculture;Washington.

(35) National Bureau of Standards (1956) "Tables of the bivariate normal distribution function and related functions. Los Angeles.

(37).....(1941) "Tables of probability functions and other tables".

(38) Maurice,Rita J. (1957) "A minimax procedure for choosing between two populations using sequential sampling". Journal Royal Statistical Society 19; 255-261.

(39) Nanda, D.N. (1949) "The standard errors of discriminant functions in plant breeding experiments". Journal Royal Statistical Society B11;283-290.

(40)Peterson, R.P. and P.G. Hoel (1949) "A solution to the problem of optimum classification". Annals of Mathematical Statistics 20; 433-438.

(41) Paulson, E. (1949) "A multiple decision procedure for certain problems in the analysis of variance". Annals Mathematical Statistics, 20; 95-98.

(42).....(1952) "On the comparison of several experimental categories with h a control". Annals Mathematical Statistics,23; 239-246.

(43).....(1952)"An optimum solution to the k sample slippage problem for the normal distribution". Annals Mathematical Statistics,23; 610-616.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(44).....(1962) "A sequential procedure for comparing several experimental categories with a control". Annals Mathematical Statistics 33; 438-443.

(45) Parzen, E. (1961) "Stochastic Processes with application to science and engineering". Prel.Edition San Francisco, Holden-Day Inc.

(46) Stein Charles (1945) "Two sample test for a linear hypothesis whose power is independent of the variance". Annals Mathematical Statistics 16; 253-258.

(47) Robertson, A. (1960) "An optimum family size in selection programmes". Biometrics 16; 296-298.

(48).....(1957) "Optimum group size in progeny testing and family selection". Biometrics 13; 442-450.

(49) Sprowls Clay (1950) "Statistical decisions by the method of minimax risk an application". Journal American Statistical Association 45;238-248.

(50) Somerville, P.N. (1954) "Some problems of optimum sampling". Biometrika 41; 420-429.

(51) Steck, G.P. (1958) "A table for computing trivariate normal probabilities". Annals Mathematical Statistics 29; 780-800.

(52)Tang, Y. (1938) "Certain Statistical problems arising in plant breeding". Biometrika 30; 29-56.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(53) Truax, Donald (1953) "An optimum slippage test for the variance of k normal distributions". Annals Math.Statistics 24; 669-674.

(54) Truax, D. and Samuel Karlin (1960) "Slippage problems". Annals Mathematical Statistics 32;296-324.

(55) Wald, A. (1950) "Statistical Decision Functions". New York; Wiley.

(56) Wilks, S.S. (1950) "Mathematical Statistics". Princeton University Press.

(57) Yates, F. (1952) "Principles governing the amount of experimentation in development work". Nature 170; 138-140.

(58) Bartlett, M.S. (1939) "The standard errors of discriminant function coefficients".Journal Royal Statistical Society, Supp.Vol.6; 169-173.

(59) Fisher, R.A. (1938) "The statistical utilization of multiple measurements". Annals Eugenics 8; 376-386.

(60) Scheffe, H. (1959) "The analysis of variance pp.78; Wiley, New York.

(61) Seal, K.C. (1955) "On a class of decision procedures for ranking means of n normal populations". Annals Math. Statistics 26; 387-398.

(62) Crammer, H. (1946) "Mathematical methods of statistics". Princeton University Press.

(63) Curnow, R.N. (1961) "Optimal programmes for varietal selection".Journal Royal Stat.Society B23;282-310.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.