5-1966

A Density for a Generalized Likelihood-Ratio Test When the Sample Size is a Random Varible

Raymond H. Neville

Utah State University

A DENSITY FOR A GENERALIZED LIKELIHOOD-RATIO TEST WHEN THE SAMPLE SIZE IS A RANDOM VARIABLE

by

Raymond H. Neville

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1966

ACKNOWLEDGMENTS

I am greatly indebted to Dr. David White, my major professor, for his assistance and encouragement in the preparation of this thesis. His constant suggestions and his readily accessible library saved many valuable hours. I extend my appreciation also to the staff of the Department of Statistics and Computer Science, Dr. Rex L. Hurst, Dr. Don Sisson, Wendell Pope, Dr. Neeti R. Bohidar, and Dr. David White for their help, encouragement, and training. I appreciate very much the corrections and suggestions given to me by the members of my graduate committee, Dr. David White, Dr. Larry O. Cannon, and Dr. Bartell Jensen.

Finally, my most sincere gratitude and appreciation are extended to my wife, Carol, for her unselfish support, devoted encouragement, and patience during our stay at this University.

Raymond H. Neville

TABLE OF CONTENTS

Page

INTRODUCTION AND REVIEW OF LITERATURE 1

One-way Classification 3

Test of Hypotheses 6

Likelihood-ratio Test 7

PROCEDURE AND RESULTS 9

Joint Density for the Observations 9

Maximum Likelihood Estimators 19

Likelihood-ratio 25

CONCLUSIONS 42

LITERATURE CITED 43

LIST OF TABLES

Table		Page
1	Symbolic representation of data in a one-way classification, N observations in i^{th} treatment	4
2	Symbolic representation of data in a one-way classification, n_i observations in i^{th} treatment	5

INTRODUCTION AND REVIEW OF LITERATURE

Many articles, books, papers, and abstracts have been published, which describe the analysis and design of experiments. R. A. Fisher's (1951) book, The Design of Experiments, is referred to by D. J. Finney (1960) as the classic for experimental designs. Some of the more outstanding books and publications in this area are those by Cochran and Cox (1957), Cox (1958), Davis et al., (1954), Federer (1955), Quenouille (1953), and Kempthorne (1952).

In most of these publications, the underlying model for the analysis, the assumptions necessary for the correct inferences, and detailed descriptions of the appropriate Analysis of Variance (A. N. O. V.) are clearly and explicitly presented. It is of interest to note, however, that the above information is based upon the assumption (although not explicitly mentioned) that the size of the samples in an experiment is a predetermined fixed quantity. That is, the experimenter, after he chooses an appropriate design, will determine the size of the samples before the experiment is actually performed. In so doing, he is also making the assumption that even though the experiment is repeated, the sample sizes will remain the same and will not vary.

The actual process of the experiment might result in lost or destroyed observations, however, making the sample sizes fluctuate or vary considerably from experiment to experiment. This variation is an indication of the randomness of the sample sizes. If there could be assigned to each possible value of the sample sizes a probability of its occurrence, then the sample sizes could be interpreted as random variables. This concept leads one to investigate the effects it might have on the A. N. O. V. for experimental designs, or on the numerous tests of hypotheses that one commonly performs, or even on the analysis of missing observations.

Although much information can be found pertaining to the analysis and design of experiments when the sample size is fixed, as indicated earlier, little has been done on this subject when sample sizes are considered to be random variables. The Statistical Theory and Method Abstracts, which give a review for all of the major statistical journals from 1959 to 1965, have been reviewed as most of the statistical literature available in the Utah State University Library, and relatively little information was found which pertained to this problem as is outlined.

Because of the apparent lack of material on this aspect of the analysis of the designs of experiments, this work will be primarily a preliminary investigation or inquiry into the effects that the assumption of random sample size might have on the tests of hypotheses in experimental designs. Also, since this is a preliminary examination, this investigation will be restricted to the simplest of designs: one-way classification.

The main objective of this work will be to examine the hypothesis that all the treatment means are the same and equal to some unknown quantity, when we know that the variance is the same for each sample, and to determine if the conventional method for making this test (the F-test) is applicable when the sample sizes are assumed to be random variables. This hypothesis can be tested by using a likelihood-ratio test. To do this, a density function or distribution has to be found for this ratio, thus permitting us to make probability statements about the occurrence of this ratio under the null hypothesis.

Throughout this development, reference will be made to many concepts of which understanding will be essential to the comprehension of the methods that have been used. Thus, a frief introduction and review will be attempted in the next few pages to prepare the reader for the material to follow. It will be assumed that the reader will have a knowledge of basic statistical terms, such as: sample, random samples, population, experimental unit, treatment, statistic, and other common, general expressions.

One-way Classification

The one-way classification, or Completely Randomized Design, is the most elementary of the experimental designs. It is represented symbolically in Tables 1 and 2. In these tables, there are t-treatments allotted at random to N' experimental units, yielding N and n_i

observations to the i^{th} treatment, respectively. Y_{ij} represents the j^{th} observation in the i^{th} treatment for each table.

Table 1. Symbolic representation of data in a one-way classification, N observations in i^{th} treatment

	Treatments						
	1	2	3		i		t
Observations	y_{11}	y_{21}	y_{31}	. . . .	y_{i1}	. . . .	y_{t1}
	y_{12}	y_{22}	y_{32}	. . . .	y_{i2}	. . . .	y_{t2}
	.	.					.
	.	.					.
	.	.					.
	.	.					.
	.	.					.
	.	.					.
	y_{1j}	y_{2j}	y_{3j}	. . . .	y_{ij}	. . . .	y_{tj}
	y_{1N}	y_{2N}	y_{3N}	. . . .	y_{iN}	. . . .	y_{tN}

Designs are usually represented by models. A model can be defined or thought of as a mathematical equation involving random variables, mathematical variables, and parameters. The distribution of the random variables, if it is known, is considered part of the model. A model for the one-way classification is given by

$$Y_{ij} = u + \tau_i + \epsilon_{ij} ,$$

where μ is the overall mean of the experiment, τ_i is the deviation of the i^{th} threatment mean $(\bar{Y}_{i.})$ from the overall mean (μ), and ϵ_{ij} is the deviation of the j^{th} observation in the i^{th} treatment (Y_{ij}) from the i^{th} treatment mean $(\bar{Y}_{i.})$ and ϵ_{ij} is normally distributed with mean zero and variance, σ_ϵ^2, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

Table 2. Symbolic representation of data in a one-way classification, n_i observations in i^{th} treatment

	Treatments						
	1	2	3		i		t
Observations	y_{11}	y_{21}	y_{31}	. . . .	y_{i1}	. . . .	y_{t1}
	y_{12}	y_{22}	y_{32}	. . . .	y_{i2}	. . . .	y_{t2}
	y_{13}	y_{23}	y_{33}	. . . .	y_{i3}		y_{t3}
	.	.	.		.		.
	.	.	.		.		.
	.	.	.		.		.
	y_{1j}	y_{2j}	y_{3j}	. . . .	y_{ij}	. . . .	y_{tj}
	.	.	.		.		.
	.	.	.		.		.
	.	.	.		.		.
	.	.	.		.		.
	.	.	.		.		.
	y_{1n_1}	y_{2n_2}	y_{3n_3}	. . . .	y_{in_j}	. . . .	y_{tn_t}

The specification of this model is not complete because nothing has been said about the τ_i. There are two possibilities: (a) the researcher can be concerned only with the t treatments in the experiment, in which case one interprets the τ_i as being a fixed effect and that $\sum_{i=1}^{t} \tau_i = 0$, or (b) he can be concerned with a population of treatments of which the t^{th} treatment is a random sample. The latter case implies that τ_i is a random effect and that $\tau_i \sim NID(0, \sigma_\tau^2)$. These two possibilities are often expressed by designating the models as Model I if τ_i is a fixed effect and as Model II if τ_i is a random effect.

One is interested when working with more than one treatment in examining various hypotheses concerning the effects of the treatments or about the populations of the treatments. Before actually making the necessary tests, certain assumptions must be made about the models. These assumptions are outlined very well in Eisenhart's (1947) papers. For the one-way classification they are: (a) the observations Y_{ij} are normally and independently distributed with mean μ_i and equal variances σ^2 ($Y_{ij} \sim NID(\mu_i, \sigma^2)$; (b) homoscedasticity; (c) $\epsilon_{ij} \sim NID(0, \sigma_\epsilon^2)$; and (d) whether the model is fixed or random. The description of the tests that are made can be given in many ways. One method used is that of the likelihood-ratio test.

Test of Hypotheses

Testing hypotheses in general involves the setting up of a hypothesis denoted H_o concerning a phenomenon in nature, and then through experimentation and sample evidence accepting or rejecting the hypothesis. It is important to note that a general hypothesis can never be proved, but can be disproved. When the experimenter takes observations and uses them as a basis for rejecting or accepting an hypothesis, he is liable to two kinds of error - the Type I error and the Type II error. The Type I error is the rejection of a hypothesis when it is true. The Type II error is the acceptance of a hypothesis when it is false. Ideally, we would like to minimize the possibility of making either of these types of errors. One usually decides on the Type I error that is permissible and then minimizes the Type II error or maximizes the power of the test. The power of the test $B(\theta)$ is defined as $B(\theta) = 1-P(II)$, where $P(II)$ is the probability of the Type II error. The power of the test $B(\theta)$ is the probability of rejecting the hypothesis when it is false. It is general practice to choose $P(I)$, the probability of making the Type I error, in advance, and then to maximize $B(\theta)$. A test which gives certain optimum properties is the likelihood-ratio test.

Likelihood-ratio Test

Define the parameter space Ω to be the set of all values that the parameters $\theta_1, \theta_2, \ldots, \theta_n$ can have and let ω denote a subspace of Ω. If we have a frequency function $f(x, \theta_1, \theta_2, \ldots, \theta_n)$, then for a sample of size n, the likelihood function is $L = \prod_{i=1}^{n} f(x_i, \theta_1, \theta_2, \ldots, \theta_n)$. If we want to test the hypothesis $H_o\left[[(\theta_1, \theta_2, \ldots, \theta_n) \epsilon \omega]\right]$ against the alternative hypothesis $H_A\left[[(\theta_1, \theta_2, \ldots, \theta_n) \epsilon \Omega - \omega]\right]$, we form the ratio

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} .$$

In the ratio above, $L(\hat{\omega})$ is the maximum of the likelihood function in the region ω with respect to the parameters that are in ω, and $L(\hat{\Omega})$ is the maximum of the likelihood function in the region Ω with respect to the parameters $\theta_1, \theta_2, \ldots, \theta_n$ that are in Ω.

λ is such that $0 \leq \lambda \leq 1$. We will reject the hypothesis if $L(\hat{\omega})$ is distant from $L(\hat{\Omega})$, and accept the hypothesis if $L(\hat{\omega})$ is close to $L(\hat{\Omega})$. We need to fix the Type I error (α) and find a constant A so that the rejection region is between 0 and A. Thus, when the hypothesis H_o is true, the Type I error, $P(I)$, will be

$$P(I) = \int_0^A g(\lambda/H_o)\, d\lambda = \alpha ,$$

where $g(\lambda/H_o)$ is the distribution for the likelihood-ratio. Thus, if λ falls in the region 0 to A, then the hypothesis H_o is rejected. If λ falls in the region A to 1, then the hypothesis is accepted.

The rest of this work will be devoted to: (a) the creation of a joint density for the observations when the sample sizes are considered to be random variables, (b) the development of a likelihood function for the joint density, and the maximum likelihood estimators for the parameters in ω and Ω, and (c) making the test $H_o(\mu_\lambda = \mu_2 = \mu_3 = \ldots = \mu_n = \mu.)$ when the sample variances are assumed to be the same unknown quantity to determine the distribution of λ. The important points will be summarized in the conclusion.

PROCEDURE AND RESULTS

Joint Density for the Observations

Consider the simplest of the experimental designs, a one-way classification, with t treatments and N experimental units per treatment, as is represented symbolically in Table 1. If we assume that there is one observation in each experimental unit, the performance of the actual experiment might result in some of the observations being lost or destroyed in some way, thus resulting in fewer experimental units per treatment and/or fewer treatments if all experimental units are lost in any one treatment. Now, it seems feasible that there could be associated with each observation a probability of its being present or absent after the experiment is performed. In other words, each experimental unit has the possibility of being lost or destroyed. This would result in the sample size associated with each treatment to vary. That is, it would become a random variable whose range of values would be from 0 to N.

Let us represent this idea symbolically. Note that in Table 1, Y_{ij} represents the j^{th} experimental unit or observation in the i^{th} treatment, and that before the experiment is performed there are N observations in each treatment. Associate with each Y_{ij} a random variable X_{ij}, which takes on the possible values 0 and 1, such that, if $X_{ij} = 0$, then Y_{ij} is absent or lost, and if $X_{ij} = 1$, then

Y_{ij} is present in the experiment. Denote the probability that X_{ij} is one, $P(X_{ij} = 1)$, by p_i and the probability that X_{ij} is zero, $P(X_{ij} = 0)$, by q_i where $q_i = 1 - p_i$ and $p_i + q_i = 1$. Let the X_{ij}'s be independent of one another. By independent is meant the occurrence of any one of the X_{ij}'s in no way affects the probability of the occurrence of the other X_{ij}'s. It is clear that each X_{ij} has two possible outcomes, zero and one. Therefore, in any trial, X_{ij} will be zero or one, and the probability density function for X_{ij} is

$$f(X_{ij}) = p_i^{X_{ij}} q_i^{1-X_{ij}} .$$

If we have N trials, then the probability that there will be n_i ones is

$$P(n_i) = \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i} .$$

Now examine an experiment with a total of N' observations, N observations per treatment with t groups or treatments. If observation Y_{ij} is present, then X_{ij} will be one. If observation Y_{ij} is absent (destroyed), then X_{ij} will be zero. Therefore, the probability that Y_{ij} is present in the experiment is

$$P(Y_{ij} \text{ is present}) = P(X_{ij} = 1) ,$$

and the probability that Y_{ij} is absent (destroyed) is

$$P(Y_{ij} \text{ is destroyed}) = P(X_{ij} = 0) .$$

Note also that the number of observations in the i^{th} treatment, after the experiment has been performed, is equal to the number of X_{ij}'s , $j = 1, 2, 3, \ldots, N$ that are equal to one, i. e., n_i = the number of observations in k^{th} treatment $= \sum_{j=1}^{N} X_{ij}$. Thus, the probability that the number of observations in the i^{th} treatment is n_i is given by

$$P(n_i \text{ observations in } i^{th} \text{ treatment}) = \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i} , \qquad \text{. . . . . (1)}$$

where $0 \le n_i \le N$.

As was mentioned in the introduction, one of the assumptions associated with the one-way classification is that the observations Y_{ij} are independent normally distributed random variables with mean μ_i and variances σ^2, $Y_{ij} \sim NID(\mu_i , \sigma^2)$. The density function for Y_{ij} is

$$f(Y_{ij}) = \begin{cases} \left(\dfrac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left[\dfrac{\left(Y_{ij} - \mu_i\right)^2}{2\sigma^2}\right] , & -\infty < Y_{ij} < \infty . \\ 0 & \text{otherwise} \end{cases}$$

With the same assumption holding for each Y_{ij}, and the added association of X_{ij} with Y_{ij}, the proposed distribution for the observations will now be conditional on X_{ij} and will be given by

$$f(Y_{ij}/X_{ij}) = \begin{cases} \left(\dfrac{1}{2\pi\sigma^2}\right)^{\frac{X_{ij}}{2}} \exp\left[\dfrac{-X_{ij}(Y_{ij}-\mu_i)^2}{2\sigma^2}\right] & \text{for } -\infty < Y_{ij} < \infty \text{ given that } X_{ij} = 1 \\ 0 & \text{for } -\infty < Y_{ij} < \infty \text{ given that } X_{ij} = 0 \\ 1 & \text{for } Y \text{ - destroyed given that } X_{ij} = 0 \\ 0 & \text{for } Y \text{ - destroyed given that } X_{ij} = 1 \end{cases} \qquad \text{. . . . . (2)}$$

In order to investigate (2) to see if it is a density function, it will be necessary to examine the sample description space associated with (2) . A sample description space S is defined as the set of all possible outcomes of an experiment. Therefore, the sample description space for (2) is

$$S = \{Y_{ij} \mid Y_{ij} \text{ is any real number or } Y_{ij} \text{ does not exist}\}$$

That is, S is composed of the set of all positive or negative numbers, and those points where Y_{ij} does not exist. S can be divided into two events. The event that Y_{ij} exists and the event that Y_{ij} does not exist. In symbolic notation,

$S = \{E_1 \cup E_2\}$, where

$E_1 = \{Y_{ij} \mid Y_{ij}$ is any real number$\}$ and

$E_2 = \{Y_{ij} \mid Y_{ij}$ does not exist$\}$.

A density must satisfy these two rules:

i) $f(Q) \geq 0$

ii) $\int_S f(Q)\, dQ = 1$.

It is obvious that rule i) is satisfied by our density, for $f(Y_{ij}/X_{ij})$ is always equal to or greater than zero when Y_{ij} is or is not present, because $f(Y_{ij}/X_{ij})$ is the normal distribution if Y_{ij} is present given that $X_{ij} = 1$ and $f(Y_{ij}/X_{ij})$ is equal to one when Y_{ij} is absent given that $X_{ij} = 0$, both of which are greater than zero. Also, $f(Y_{ij}/X_{ij}) = 0$, when Y_{ij} is present given that $X_{ij} = 0$, and when Y_{ij} is absent given that $X_{ij} = 1$. (2) also satisfies rule ii), since

$$\int_S f(Y_{ij}/X_{ij})dY_{ij} = \int_{E_1} f(Y_{ij}/X_{ij})dY_{ij} + \int_{E_2} f(Y_{ij}/X_{ij})dY_{ij} \quad . \qquad (3)$$

When $X_{ij} = 0$ or $X_{ij} = 1$, (3) equals one. This is obvious since

$$\int_{E_1} f(Y_{ij}/X_{ij} = 0)\, dY_{ij} + \int_{E_2} f(Y_{ij}/X_{ij} = 0)\, dY_{ij} = 1 \;,$$

and

$$\int_{E_1} f(Y_{ij}/X_{ij} = 1)\, dY_{ij} + \int_{E_2} f(Y_{ij}/X_{ij} = 1)\, dY_{ij} = 1 \;,$$

for

$$\int_{E_1} f(Y_{ij}/X_{ij} = 0)\, dY_{ij} = \int_{E_2} f(Y_{ij}/X_{ij} = 1)\, dY_{ij} = 0$$

and

$$\int_{E_1} f(Y_{ij}/X_{ij} = 1)\, dY_{ij} = \int_{E_2} f(Y_{ij}/X_{ij} = 0)\, dY_{ij} = 1 \;.$$

therefore, (3) is always one which implies that (2) is a density function. The joint density of Y_{ij} and X_{ij} is

$$g(Y_{ij}\,,\, X_{ij}) = f(Y_{ij}/X_{ij})\; P(X_{ij}) \;.$$

That is,

$$g(Y_{ij}, X_{ij}) = \begin{cases} 0 & \text{for } -\infty < Y_{ij} < \infty,\ X_{ij} = 0 \text{ and } 0 < p_i < 1 \\ \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{X_{ij}}{2}} \exp\left[\frac{-X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right] . q_i & \text{for } -\infty < Y_{ij} < \infty,\ X_{ij} = 1, \\ & \text{and } 0 < q_i < 1 \\ 0 & \text{for } Y \text{ destroyed, } X_{ij} = 1,\ 0 < p_i < 1 \\ p_i & \text{for } Y \text{ destroyed, } X_{ij} = 0,\ 0 < p_i < 1 \end{cases}$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The Sample Description Space S for this joint distribution is:

$$S = \{(Y_{ij}, X_{ij}) \mid -\infty < Y_{ij} < \infty,\ Y_{ij} \text{ is destroyed, } X_{ij} = 0, 1\} .$$

S can be partitioned into four subsets or events:

$$S = \left(E_1 \cup E_2 \cup E_3 \cup E_4\right) \text{ where}$$

$$E_1 = \left((Y_{ij}, X_{ij}) \mid (Y_{ij} \text{ is any real number and } X_{ij} = 0)\right),$$

$$E_2 = \left((Y_{ij}, X_{ij}) \mid (Y_{ij} \text{ is any real number and } X_{ij} = 1)\right),$$

$$E_3 = \left((Y_{ij}, X_{ij}) \mid (Y_{ij} \text{ does not exist and } X_{ij} = 0)\right), \text{ and}$$

$$E_4 = \left((Y_{ij}, X_{ij}) \mid (Y_{ij} \text{ does not exist and } X_{ij} = 1)\right).$$

It is obvious from (4) that $f(Y_{ij}, X_{ij}) \geq 0$ and therefore, rule i) is satisfied. Rule ii) will also be satisfied, since,

$$\int_S \sum_{X_{ij}=0}^{1} q(Y_{ij}, X_{ij})dY_{ij} = \int_S \sum_{X_{ij}=0}^{1} f(Y_{ij}/X_{ij}) \cdot P(X_{ij})dY_{ij}$$

$$= \int_S \left(f(Y_{ij}/X_{ij} = 0) \cdot q_i + f(Y_{ij}/X_{ij} = 1) \cdot p_i \right) dY_{ij}$$

$$= q_i \left(\int_{E_1} f(Y_{ij}/X_{ij} = 0)dY_{ij} + \int_{E_2} f(Y_{ij}/X_{ij} = 0)dY_{ij} \right)$$

$$+ p_i \left(\int_{E_1} f(Y_{ij}/X_{ij} = 1)dY_{ij} + \int_{E_2} f(Y_{ij}/X_{ij} = 1)dY_{ij} \right)$$

$$= q_i + p_i = 1 .$$

Therefore, (4) is a density function.

It can be shown, Parzen (1960), and Feller (1957), that if A and B are independent random variables, then their joint density can be obtained by the product of their respective density functions. That is,

$$f(A, B) = f(A)f(B) \qquad \qquad (5)$$

Since the Y_{ij}'s, $j = 1, 2, \ldots, N$, are independent normally distributed random variables, their joint density given the corresponding X_{ij}'s, $j = 1, 2, \ldots, N$, is

$$f(Y_{i1}, Y_{i2}, \ldots, Y_{iN}/X_{i1}, X_{i2}, \ldots, X_{iN})$$

$$= \prod_{j=1}^{N} f(Y_{ij}/X_{ij}) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2 \sum_{j=1}^{N} X_{ij}}$$

$$\exp\left(-\sum_{j=1}^{N} \frac{X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right) \quad \text{for } -\infty < Y_{ij} < \infty \text{ given } X_{ij} = 1,$$

$$Y_{ij} \text{ destroyed given } X_{ij} = 0;$$

$$= 0 \qquad \text{otherwise}.$$

The joint density for the Y_{ij} and X_{ij}, $j = 1, \ldots, N$, is

$$f(Y_{i1}, \ldots, Y_{iN}, X_{i1}, \ldots, X_{iN}) = f(Y_{i1}, \ldots, Y_{iN}/X_{i1}, \ldots,$$

$$X_{iN}) \;.\; P(X_{i1}, \ldots, X_{iN})$$

or,

$$f(Y_{i1}, \ldots, Y_{iN}, X_{i1}, \ldots, X_{iN}) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2 \sum_{j=1}^{N} X_{ij}}$$

$$\exp\left(-\sum_{j=1}^{N} \frac{X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right) \cdot \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i}$$

for $-\infty < Y_{ij} < \infty$, and $X_{ij} = 1$, Y_{ij} destroyed and

$X_{ij} = 0$, $0 \leq p_i$, $q_i \leq 1$, $0 \leq n_i \leq N$, and 0 otherwise.

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{6}$$

This is a density function because each density $f(Y_{ij}, X_{ij})$ is greater than or equal to zero for all values of Y_{ij} and X_{ij}, thus, the product of positive quantities will be positive, and

$$\int_S \sum_{X_{ij}=0} f(Y_{ij}, X_{ij}) dY_{ij} = 1 .$$

Therefore,

$$\int_{S_1} \int_{S_2} \cdots \int_{S_N} \sum_{X_{i1}=0}^{1} \sum_{X_{i2}=0}^{1} \cdots \sum_{X_{iN}=0}^{1} \prod_{j=1}^{N} f(Y_{ij}, X_{ij}) dY_{ij}$$

$$= \prod_{j=1}^{N} \int_{S_j} \sum_{X_{ij}=0}^{1} f(Y_{ij}, X_{ij}) dY_{ij} = 1 ,$$

and rule ii) is satisfied.

The joint distribution of the Y_{ij}'s and X_{ij}'s for all of the t treatments is given by

$$f(Y_{11}, Y_{12}, \ldots, Y_{1N}, Y_{21}, \ldots, Y_{2N}, \ldots, Y_{tN}, X_{11}, \ldots,$$

$$X_{1N}, \ldots, X_{tN}) = \prod_{i=1}^{t} f(Y_{i1}, \ldots, Y_{iN}/X_{i1}, \ldots, X_{iN})$$

$$P(X_{i1}, \ldots, X_{iN})$$

$$= \prod_{i=1}^{t} \left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n_i}{2}} \exp\left[\frac{-\sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right] \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i} \right] \quad \ldots\ldots\ldots\ldots \quad (7)$$

for $-\infty < Y_{ij} < \infty$ and $X_{ij} = 1$, Y_{ij} destroyed and $X_{ij} = 0$, $0 \leq n_i \leq N$, $1 \geq p_i, q_i \geq 0$, $p_i + q_i = 1$, and 0 otherwise. By the same reasoning as that used previously, it can be shown that (7) is also a density function.

Now that we have the joint density function, we are in a position to test the hypothesis, H_o: $(\mu_1 = \mu_2 \ldots = \mu_t = \mu_0)$. As was mentioned in the introduction, we can make the test by the likelihood-ratio criterion; however, to do this we need the maximum likelihood estimators for the unknown parameters. This, then, will be the next topic discussed.

Maximum Likelihood Estimators

As was indicated in the introduction, to find the maximum likelihood estimators of the unknown parameters, we maximize the likelihood functions with respect to the unknown parameters over the regions in which these parameters are defined. We, in essence, have two regions, ω and Ω.

The Likelihood Function for the Region ω

The subspace ω is defined as follows:

$$\omega = \{\mu, \sigma^2, p_i \mid \mu = \mu_O, \ \sigma^2 > 0, \ 0 \le p_i \le 1 \} .$$

Therefore, the likelihood function defined on the region ω is

$$L(\mu, \sigma^2, p_i) = \prod_{i=1}^{t} \left(\left[\left(\frac{1}{2\pi\sigma^2}\right)^{n_i/2} \exp\left(\sum_{j=1}^{N} \frac{X_{ij}(Y_{ij} - \mu_O)^2}{2\sigma^2} \right) \right] \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i} \right)$$

or

$$L(\mu, \sigma^2, p_i) = \left(\frac{1}{2\pi\sigma^2}\right)^{\sum_{i=1}^{t} n_i/2} \exp\left[\frac{-\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_O)^2}{2\sigma^2} \right] \prod_{i=1}^{t} \binom{N}{n_i} p_i^{n_i} q_i^{N-n_i}$$

The natural log of this function gives

$$L^*(\mu_O, \sigma^2, p_i) = -\sum_{i=1}^{t} n_i/2 \ \ln(2\pi\sigma^2) \ \frac{-\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_O)^2}{2\sigma^2}$$

$$+ \sum_{i=1}^{t} \left[\ln \binom{N}{n_i} + n_i \ln(p_i) + (N - n_i) \ln(q_i) \right].$$

Now, we want to find the estimates of the parameters which will maximize $L(\mu_o, \sigma^2, p_i)$. To do this, we need to take the partial derivatives of $L(\mu_o, \sigma^2, p_i)$ with respect to the parameters μ_o, σ^2, and p_i, and by setting these first order derivatives to zero solve for μ_o, σ^2, and p_i, if the values exist.

$$\frac{\partial L^*(\mu_o, \sigma^2, p_i)}{\partial \mu_o} = -2 \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_o)^2}{2\sigma^2} (-1)$$

Setting this quantity to zero implies that

$$\hat{\mu}_o = \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij} Y_{ij}}{\sum_{i=1}^{t} n_i} = \bar{Y}_{..} .$$

$$\frac{\partial L^*(\mu_o, \sigma^2, p)}{\partial \sigma^2} = -\sum_{i=1}^{t} n_i \frac{1}{\sigma^2} + \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{..})^2}{2(\sigma^2)^2} .$$

Setting this quantity to zero implies that

$$-\sum_{i=1}^{t} n_i/2 \cdot \frac{\sigma^2}{\sigma^4} + \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{..})^2}{2\sigma^4} = 0 .$$

This implies that

$$\hat{\sigma}^2_\omega = \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{..})^2}{\sum_{i=1}^{t} n_i}$$

$$\frac{\partial L^*(\mu_o, \sigma^2, p_i)}{\partial p_i} = \frac{n_i}{p_i} + \frac{N-n_i}{q_i} \quad (-1)$$

Setting this quantity to zero implies that

$$\frac{n_i}{p_i} - \frac{(N-n_i)}{(1-p_i)} = \frac{n_i - n_i p_i - p_i N + n_i p_i}{p_i(1-p_i)} = 0 \ .$$

This implies that

$$\hat{p}_i = \frac{n_i}{N}$$

The parameter space (Ω) is defined as the set of all values of $\mu_1, \mu_2, \ldots, \mu_t, \sigma^2$, and p_i, i=1, 111, t, such that $\mu_i \neq \mu_o, \sigma^2 > 0$, and $1 \geq p_i \geq 0$. Symbolically, $\Omega = \{\mu_i, p_i, i=1, \ldots, t, \sigma^2 | \mu_i \neq \mu_o, \sigma^2 > 0, 0 \leq p_i \leq 1\}$. The likelihood function defined on the parameter space Ω is

$$L(\mu_i, \mu_2, \ldots, \mu_t, \sigma^2, p_i) = \prod_{i=1}^{t} \left[\left(\frac{1}{2\pi\sigma^2}\right)^{n_i/2} \right.$$

$$\exp\left[\left\{\frac{-\sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right\} \cdot \binom{N}{n_i} p_i^{\,n_i} \; q_i^{\,(N-n_i)}\right],$$

or

$$L(\mu_i, \ldots, \mu_t, {}^2, p_i) = \left(\frac{1}{2\sigma^2}\right)^{\sum_{i=1}^{t} n_i/2}$$

$$\exp\left[\frac{-\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2}\right] \cdot P(n_1, \ldots, n_t) .$$

The natural log of this function gives

$$L^*(\mu_1, \ldots, \mu_t, \sigma^2, p_i) = \frac{-\sum_{i=1}^{t} n_i}{2} \ln(2\pi\sigma^2)$$

$$\frac{-\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_i)^2}{2\sigma^2} + \ln\Big(P(n_1, n_2, \ldots, n_t)\Big) .$$

The maximum likelihood estimators are obtained as follows:

$$\frac{\partial L^*(\mu_1, \ldots, \mu_t, \sigma^2, p_i)}{\partial \mu_i} = -2 \, \frac{\sum_{j=1}^{N} X_{ij}(Y_{ij} - \mu_i)^2 \, (-1)}{2\sigma^2}$$

Setting this quantity to zero implies that

$$\hat{\mu}_i = \frac{\sum_{j=1}^{N} X_{ij} Y_{ij}}{\sum_{j=1}^{N} X_{ij}} = \bar{Y}_{i.}\ , \qquad i = 1, 2, \ldots, t\ ,$$

and

$$\frac{\partial L^*(\mu_1, \ldots, \mu_t, \sigma^2, p_i)}{\partial \sigma^2} = -\frac{\sum_{i=1}^{t} n_i}{2} \frac{1}{\sigma^2}$$

$$+ \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - Y_{ij} - \bar{Y}_{i.})^2}{2\sigma^4}$$

Setting this quantity to zero implies that

$$-\frac{\sum_{i=1}^{t} n_i \sigma^2}{2\sigma^4} + \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2}{2\sigma^4} = 0$$

$$\hat{\sigma}_{\Omega} = \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2}{\sum_{i=1}^{t} n_i}$$

and

$$\hat{p}_{i\Omega} = \frac{n_i}{N}$$

Likelihood-ratio

It was indicated in the introduction that to test the hypothesis $H_o(\theta \epsilon\, \omega)$ against the hypothesis $H_A(\theta \epsilon\, \Omega - \omega)$ we need to calculate the following ratio

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{L(\hat{\mu}_o, \hat{\sigma}^2_\omega)}{L(\hat{\mu}_1, \ldots, \hat{\mu}_t, \hat{\sigma}^2_\Omega)}$$

For our problem, the likelihood-ratio is

$$\lambda = \frac{\left(\frac{1}{2\pi\hat{\sigma}^2_\omega}\right)^{\sum_{i=1}^{t} n_i/2} \exp\left(-\frac{\sum_{i=1}^{t}\sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{..})^2}{2\hat{\sigma}^2_\omega}\right) P(n_1, n_2, \ldots, n_t)}{\left(\frac{1}{2\pi\hat{\sigma}^2_\Omega}\right)^{\sum_{i=1}^{t} n_i/2} \exp\left(-\frac{\sum_{i=1}^{t}\sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2}{2\hat{\sigma}^2_\Omega}\right) P(n_1, \ldots, n_t)}$$

but in ω

$$\hat{\sigma}^2_\omega = \frac{\sum_{i=1}^{t}\sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{..})^2}{\sum_{i=1}^{t} n_i}$$

and in Ω

$$\hat{\sigma}^2_{\Omega} = \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \overline{Y}_{i.})^2}{\sum_{i=1}^{t} n_i} .$$

Substituting equals for equals and cancelling like terms, the likelihood-ratio becomes

$$\lambda = \left[\frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \overline{Y}_{i.})^2}{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \overline{Y}_{..})^2} \right]^{\frac{\sum_{i=1}^{t} n_i}{2}} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (8)$$

It can be shown quite easily that

$$\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \overline{Y}_{i.})^2$$

$$+ \sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(\overline{Y}_{i.} - \overline{Y}_{..})^2 \;;$$

hence, by substituting this relationship into the expression above, (8) becomes

$$\lambda = \left[\frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2}{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2 + \sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(\bar{Y}_{i.} - \bar{Y}_{..})^2} \right]^{\sum_{i=1}^{t} n_i/2} \qquad \ldots\ldots\ldots\ldots (9)$$

$$= \left[\frac{1}{1 + \dfrac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2}} \right]^{\sum_{i=1}^{t} n_i/2} \qquad \ldots\ldots\ldots\ldots (10)$$

Note 1

$$V = \frac{\sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij}(\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sigma^2}$$

is distributed as a Chi-square variate with $\ell - 1$ degrees of freedom, where

$$\ell = \sum_{i=1}^{t} \left[1 - \sum_{k=0}^{n_i} \binom{n_i}{k} (-1)^k \right] .$$

The relationship for ℓ is obtained by noting that if all the observations in any one treatment are missing, then likewise so is the treatment. This is represented symbolically in this manner:

$$X = \begin{cases} 0 & \text{if } ij^{th} \text{ observation is missing} \\ 1 & \text{if } ij^{th} \text{ observation is present} \end{cases}$$

$$\prod_{j=1}^{N} (1 - X_{ij}) = \begin{cases} 0 & \text{if at least one observation in } i^{th} \text{ treatment is present} \\ 1 & \text{if all observations in } i^{th} \text{ treatment are missing} \end{cases} \quad \ldots\ldots\ldots\ldots \quad (11)$$

$$1 - \prod_{j=1}^{N} (1 - X_{ij}) = \begin{cases} 0 & \text{if all observations in } i^{th} \text{ treatment are missing} \\ 1 & \text{if at least one observation is present in } i^{th} \text{ treatment} \end{cases}$$

$$\ldots\ldots\ldots\ldots \quad (12)$$

$\prod_{j=1}^{N} (1 - X_{ij})$ can be represented by $\sum_{k=0}^{n_i} (-1)^k \binom{n_i}{k}$, by noting that

$$(X - Y)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} X^k Y^{n-k}$$

Let $X = Y = 1$, and define $0^0 = 1$, then

$$\prod_{j=1}^{N} (1 - X_{ij}) = (1 - 1)^{n_i} = 0^{n_i} = \begin{cases} 0 & \text{if } n_i > 0 \\ 1 & \text{if } n_i = 0 \end{cases}$$

This implies that if $X = Y = 1$, then

$$(X - Y)^{n_i} = \sum_{k=0}^{n_i} (-1)^k \binom{n_i}{k} ;$$

hence,

$$\ell = \sum_{i=1}^{t} (1 - \prod_{j=1}^{N} (1 - X_{ij})) = \sum_{i=1}^{t} (1 - \sum_{k=0}^{n_i} (-1)^k \binom{n_i}{k}) .$$

The range for ℓ is 0 to t .

Note 2

$$U = \sum_{i=1}^{t} \sum_{j=1}^{N} X_{ij} (Y_{ij} - Y_{i.})^2$$

is distributed as a Chi-square variate with $\sum_{i=1}^{t} (n_i - 1)$ degrees of freedom.

Note 3

$\dfrac{\frac{V}{m}}{\frac{U}{n}}$, where $m = \ell - 1$ and $n = \sum_{i=1}^{t} (n_i - 1)$ are the degrees of freedom associated with the Chi-square variates V and U, respectively, is distributed as an F variate with m and n degrees of freedom. This relationship is conditional on holding the n_i's constant.

Note 4

The distribution for F, Mood and Graybill (1963), is given by

$$h(F) = \frac{\left(\frac{m+n-2}{2}\right)!}{\left(\frac{m-2}{2}\right)!\left(\frac{n-2}{2}\right)!}\left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{F^{\frac{m-2}{2}}}{\left(1+\frac{m}{n}F\right)^{\frac{m+n}{2}}}$$

for $F > 0$.

Using the information in the above notes, we see that

$$\frac{\dfrac{\sum_{i=1}^{t}\sum_{j=1}^{N} X_{ij}(\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sigma^2(\ell - 1)}(\ell - 1)}{\dfrac{\sum_{i=1}^{t}\sum_{j=1}^{N} X_{ij}(Y_{ij} - \bar{Y}_{i.})^2 \sum_{i=1}^{t}(n_i - 1)}{\sigma^2\left(\sum_{i=1}^{t}(n_i - 1)\right)}}$$

is distributed as an F variate with $m = \ell - 1$ and $n = \sum_{i=1}^{t}(n_i - 1)$ degrees of freedom, and therefore (10) becomes

$$\lambda = \left[\frac{1}{1 + \left[\dfrac{\ell - 1}{\sum_{i=1}^{t}(n_i - 1)}\right] F}\right]^{\frac{\sum_{i=1}^{t} n_i}{2}} \qquad \text{. . . . . (13)}$$

By a simple transformation, Parzen (1960), and Mood and Graybill (1963), of variables and letting $m = \ell - 1$, $Z = \Sigma n_i$, and $n = \sum_{i=1}^{t} (n_i - 1)$, we can find the distribution of λ as follows:

$$\lambda = \left(\frac{1}{1 + \frac{m}{n} F} \right)^{Z/2}$$

Solving for F, gives

$$\lambda^{2/Z} = \left(\frac{1}{1 + \frac{m}{n} F} \right)$$

$$\lambda^{2/Z} \left(1 + \frac{m}{n} F\right) = 1$$

$$\frac{m}{n} F = \left(\lambda^{-2/Z} - 1 \right)$$

$$F = \left(\lambda^{-2/Z} - 1 \right) (n/m)$$

$$\left| \frac{dF}{d\lambda} \right| = \left[(2/Z) \lambda^{-(2/Z) - 1} \right] \left(\frac{n}{m} \right)$$

$$h(\lambda) = \frac{\left(\frac{m+n-2}{2}\right)!}{\left(\frac{m-2}{2}\right)!\left(\frac{n-2}{2}\right)!}\left(\frac{m}{n}\right)^{m/2} \frac{\left(\frac{1-\lambda^{2/Z}}{\lambda^{2/Z}}\right)^{\frac{m-2}{2}} \left(\frac{m}{n}\right)^{\frac{m-2}{2}}}{\left[1+\frac{m}{n}\cdot\frac{n}{m}\,\frac{1-\lambda^{2/Z}}{\lambda^{2/Z}}\right]^{\frac{m+n}{2}}} \cdot$$

$$\left(\frac{n}{m}\right)\left(\frac{2}{Z}\right)\frac{1}{\lambda^{(\,(2/Z)+1)}}$$

$$= C \cdot \left(\frac{m}{n}\right)^{m/2} \cdot \left(\frac{n}{m}\right)^{m/2} \left(\frac{2}{Z}\right) \frac{\left[\frac{1-\lambda^{2/Z}}{\lambda^{2/Z}}\right]^{\frac{m-2}{2}}}{\left[\frac{\lambda^{2/Z}+1-\lambda^{2/Z}}{\lambda^{2/Z}}\right]^{\frac{m+n}{2}}} \cdot \frac{1}{\lambda^{(2/Z+1)}}$$

$$= C \cdot 2/Z \left[\frac{1-\lambda^{2/Z}}{\lambda^{2/Z}}\right]^{\frac{m-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{\frac{m+n}{2}} \cdot \left(\frac{1}{\lambda^{(2/Z+1)}}\right)$$

$$= C \cdot 2/Z \frac{\left(1-\lambda^{2/Z}\right)^{\frac{m-2}{2}}}{\left(\lambda^{2/Z}\right)^{\frac{m-2}{2}}} \cdot \left[\lambda^{2/Z}\right]^{\frac{m+n}{2}} \cdot \lambda^{-(2/Z+1)}$$

$$= C \cdot 2/Z \left[1-\lambda^{2/Z}\right]^{\frac{m-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{\frac{m+n}{2}-\frac{m-2}{2}} \cdot \lambda^{-(2/Z+1)}$$

$$= C \cdot 2/Z \left[1-\lambda^{2/Z}\right]^{\frac{m-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{\frac{m+n-m+2}{2}} \cdot \left[\lambda^{2/Z}\right]^{-1} \cdot \lambda^{\frac{-2Z}{2Z}}$$

$$= C \cdot 2/Z \left[1-\lambda^{2/Z}\right]^{\frac{m-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{\frac{n+2-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{-Z/2}$$

$$= C \cdot 2/Z \left[1-\lambda^{2/Z}\right]^{\frac{m-2}{2}} \cdot \left[\lambda^{2/Z}\right]^{\frac{n-Z}{2}}$$

where $m = \ell - 1$, $n = \sum_{i=1}^{t} \left[\sum_{j=1}^{N} X_{ij} - 1\right]$, $Z = \sum_{i=1}^{t} \sum_{j=1}^{n} X_{ij}$, and

$C = \dfrac{\left(\frac{m+n-2}{2}\right)!}{\left(\frac{m-2}{2}\right)!\left(\frac{n-2}{2}\right)!}$. Making the substitutions mentioned above

$h(\lambda)$ becomes

$$h(\lambda) = \frac{\left[\frac{\Sigma\Sigma X_{ij} - 3}{2}\right]!}{\left(\frac{\ell-3}{2}\right)!\left[\frac{\Sigma\Sigma X_{ij} - \ell - 2}{2}\right]!} \left[\frac{2}{\Sigma\Sigma X_{ij}}\right] \left[1-\lambda^{2/\Sigma\Sigma X_{ij}}\right]^{\frac{\ell-3}{2}} \left[\lambda^{2/\Sigma\Sigma X_{ij}}\right]^{\frac{-\ell}{2}}$$

. . . . (14)

and in terms of $n_i = \sum_{j=1}^{N} X_{ij}$

$$h(\lambda) = \frac{\left[\frac{\Sigma n_i - 3}{2}\right]!}{\left(\frac{\ell-3}{2}\right)!\left[\frac{\Sigma n_i - \ell - 2}{2}\right]!} \left[\frac{2}{\Sigma n_i}\right] \left[(1-\lambda)^{\left(2/\Sigma n_i\right)}\right]^{\frac{\ell-3}{2}} \left[\lambda^{\left(2/\Sigma n_i\right)}\right]^{\frac{-\ell}{2}}$$

. . . . (1 5)

where $0 \leq \lambda \leq 1$.

A preliminary investigation of the distribution for λ that has just been obtained, reveals that this distribution depends upon the values of n_i, $i = 1, 2, \ldots, t$, and is therefore conditional on n_i, $i = 1, 2, \ldots, t$.

This can be represented as

$$h(\lambda / n_i, n_2, \ldots, n_t) = \frac{\left[\frac{\Sigma n_i - 3}{2}\right]!}{\left(\frac{\ell - 3}{2}\right)! \left[\frac{\Sigma n_i - \ell - 2}{2}\right]!} \left[\frac{2}{\Sigma n_i}\right]$$

$$\left[(1-\lambda)^{\left(2/\Sigma n_i\right)}\right]^{\frac{\ell - 3}{2}} \left[\lambda^{\left(2/ \quad n_i\right)}\right]^{\frac{-\ell}{2}} \qquad 0 < \lambda < 1 .$$

$$\ldots \ldots \ldots \ldots \qquad (16)$$

The joint dnesity for λ and $n_1, n_2, \ldots, n_t$ is

$$h(\lambda, n_1, n_2, \ldots, n_t) = h(\lambda / n_1, n_2, \ldots, n_t) \cdot P(n_1, n_2, \ldots, n_t)$$

$$\ldots \ldots \ldots \ldots \qquad (17)$$

where

$$P(n_1, n_2, \ldots, n_t) = \prod_{i=1}^{t} \binom{N}{n_i} P_i^{\,n_i} \; q_i^{\,N - n_i} .$$

The distribution for λ is found by summing (17) over all possible values for the n_i's, $i=1, 2, \ldots, t$. That is

$$h(\lambda) = \sum_{n_i=0}^{N} \cdots\cdots \sum_{n_t=0}^{N} h(\lambda/n_1, \ldots, n_t) \cdot P(n_1, \ldots, n_t) \,. \qquad \ldots\ldots\ldots\ldots \quad (18)$$

To simplify the form of this distribution, make the substitution

$$W = \lambda^{2/\Sigma n_i} \,.$$

This implies that

$$\lambda = W^{\Sigma n_i/2} \quad \text{and} \quad \left|\frac{d\lambda}{dW}\right| = \frac{\sum_{i=1}^{t} n_i}{2} \, W^{(\Sigma n_i/2 - 1)} \,.$$

Making the appropriate substitutions,

$$h(W/n_i, n_2, \ldots, n_t) = \frac{\left[\frac{\Sigma n_i - 3}{2}\right]!}{\left(\frac{\ell-3}{2}\right)! \left[\frac{\Sigma n_i - \ell - 2}{2}\right]!} \left[\frac{2}{\Sigma n_i}\right] \cdot \left[\frac{\Sigma n_i}{2}\right]$$

$$\cdot \, (1 - W)^{\frac{\ell-3}{2}} \cdot W^{-\ell/2} \cdot W^{(\Sigma n_i/2 - 1)}$$

$$= \frac{\left[\frac{\Sigma n_i - 3}{2}\right]!}{\left(\frac{\ell-3}{2}\right)! \left[\frac{\Sigma n_i - \ell - 2}{2}\right]!} (1-W)^{\frac{\ell-3}{2}} \, W^{\left(\frac{\Sigma n_i - \ell - 2}{2}\right)} ; \quad 0 < W < 1 \,.$$

$$\ldots\ldots\ldots\ldots \quad (19)$$

The joint distribution for W and n_1, $i=1, \ldots, t$ is

$h(W, n_1, \ldots, n_t) = h(W/n_1, \ldots, n_t) \cdot P(n_1, \ldots, n_t)$, or

$$h(W, n_1, \ldots, n_t) = \frac{\left(\frac{\Sigma n_i - 3}{2}\right)!}{\left(\frac{\ell-3}{2}\right)!\left(\frac{\Sigma n_i - \ell - 2}{2}\right)!} (1 - W)^{\frac{\ell-3}{2}}$$

$$W^{\left(\frac{\Sigma n_i - \ell - 2}{2}\right)} \cdot \prod_{i=1}^{t} \left[\binom{N}{n_i} p_i^{n_i} q_i^{N-n_i}\right], \quad 0 < W < 1, \ 0 \leq p_i \leq 1 .$$

$$\ldots \ldots \ldots \ldots \quad (20)$$

Summing over all possible values of $n_1, \ldots, n_t$ will give the density for W, i. e.,

$$g(W) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} \cdots \sum_{n_t=0}^{N} h(W, n_1, \ldots, n_t)$$

$$\ldots \ldots \ldots \ldots \quad (21)$$

This, then gives us a density function for W. To complete the test of hypothesis, one needs to find the critical region, $0 - A$, such that $P(0 < W < A) = \alpha$, where α is the Type I error probability and is usually chosen beforehand.

It should be pointed out that

$$P(0 < W < A) = \int_0^A g(W)dW = \int_0^A \sum_{n_i=0}^{N} \cdots \sum_{n_t=0}^{N} h(W, n_1, \ldots, n_t)dW$$

and

$$P(0 < \lambda < A) = \int_0^A h(\lambda)d\lambda = \int_0^A \sum_{n_i=0}^{N} \cdots \sum_{n_t=0}^{N} h(\lambda, n_i, \ldots, n_t)d\lambda$$

are not immediately obtainable, and that a test for λ or W will not be obtained from these relationships until a closed expression can be found for $g(W)$ and $h(\lambda)$ or until an approximation can be obtained for them which will permit the calculation of the probabilities of (18) and (21). Even if we were able to obtain (18) and (21) a test still might be difficult to obtain because both (18) and (21) involve the unknown parameters p_i and q_i. The Type I error would vary for the different values of the parameters for a fixed critical region 0 to A.

Mood and Graybill (1963) indicate a method, however, which will permit us to construct a test. They point out that if a test criterion λ has a distribution $f(\lambda; \theta_1, \theta_2, \ldots, \theta_n)$ which involves a set of unknown parameters $\theta_1, \theta_2, \ldots, \theta_n$, and these parameters have a set of sufficient statistics $\hat{\theta}_1, \ldots, \hat{\theta}_n$, then the joint density for λ and $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ may be expressed as

$$v(\lambda, \hat{\theta}_1, \ldots, \hat{\theta}_n; \theta_1, \ldots, \theta_n) = v_1(\lambda/\hat{\theta}_1, \ldots, \hat{\theta}_n)$$

$$v_2(\hat{\theta}_1, \ldots, \hat{\theta}_n; \theta_1, \ldots, \theta_n) .$$

$$\ldots \ldots \ldots \quad (22)$$

A sufficient statistic implies that the estimator contains all the information about the true parameter that the sample can give. From (22) one notes that the conditional density of λ, given the sufficient statistics, will not involve the parameters. Using this conditional distribution, a number $A(\hat{\theta}_1, \ldots, \hat{\theta}_n)$ may be found which for every $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$

$$\int_0^{A(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)} v_1(\lambda/\hat{\theta}_1, \ldots, \hat{\theta}_n)\, d\lambda = \alpha$$

is true. Hence, one may test a hypothesis by using $\hat{\theta}_1, \ldots, \hat{\theta}_n$ and λ. The test is actually a conditional test. We can observe $\hat{\theta}_1, \ldots, \hat{\theta}_n$ and test λ by the critical region $0 < \lambda < A(\theta_1, \theta_n)$, using the conditional distribution of λ given $\hat{\theta}_1, \ldots, \hat{\theta}_n$.

The important point to be gleaned from the above discussion is that if a set of sufficient statistics exist, then

$$\int_0^{A(\hat{\theta}_1, \ldots, \hat{\theta}_n)} v_1(\lambda/\hat{\theta}_1, \ldots, \hat{\theta}_n)\, d\lambda = \alpha$$

will give a test for the likelihood criterion λ for arbitrary values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$.

There is a theorem in Mood and Graybill (1963) that gives a criterion for examining a set of statistics for sufficiency. This theorem states that if a joint dnesity of a random sample can be factored as

$$g(X_1, \ldots, X_n; \theta_1, \ldots, \theta_n) = h(\hat{\theta}_1, \ldots, \hat{\theta}_n; \theta_1, \ldots, \theta_n) \cdot$$

$g(X_1, \ldots, X_n)$ where $g(X_1, \ldots, X_n)$ does not involve the θ_i, then $\hat{\theta}_1, \ldots, \hat{\theta}_n$ is a set of n sufficient statistics.

Now, the joint densities for λ or W, which were just obtai ned in (17) and (20), respectively, involve the unknown parameters $p_1, p_2, \ldots, p_t$. Note that (17) and (20), represented symbolically below,

$$h(\lambda, n_1, \ldots, n_t; p_1, \ldots, p_t) = g(\lambda/n_1, \ldots, n_t)\, f(n_1, \ldots, n_t; p_i, p_2, p_t)$$

$$h(W, n_1, \ldots, n_t; p_1, \ldots, p_t) = g(W/n_1, \ldots, n_t)\, f(n_i, \ldots, n_t;$$
$$p_1, \ldots, p_t)\ ,$$

are the product of two densities where the first density does not involve the parameters $p_1, \ldots, p_t$, and the second density involves n_i and p_i, $i=1, \ldots, t$.

Using the criterion for sufficiency outlined previously, one sees that the n_i's, $i=1, \ldots, t$ are sufficient statistics for the parameters p_i, $i=1, \ldots, t$. Therefore, a test can be performed on λ or W by using

$$\int_0^{A(n_1, \ldots, n_t)} g(\lambda / n_1, \ldots, n_t)\, d\lambda = \alpha \qquad \ldots\ldots\ldots\ldots\ldots \quad (23)$$

or

$$\int_0^{A(n_1, \ldots, n_t)} g(W / n_1, \ldots, n_t)\, dW = \alpha \qquad \ldots\ldots\ldots\ldots\ldots \quad (24)$$

It is possible to demonstrate this by noting that to test the hypothesis $H_o(\mu_1 = \mu_2 = \ldots = \mu_t = \mu_0)$ we need to find a critical region for λ or W, such that if any value of λ or W obtained by the likelihood-ratio falls in this region, then H_o will be rejected. If the value for λ or W is not in this region then H_o is accepted. This is expressed symbolically as

$$\sum_{n_i} \int_0^{A(n_i)} g(s^*, n_i; p_i) ds^*$$

$$= \sum_{n_i} \int_0^{A(n_i)} g(s^*/n_i) f(n_i; p_i) ds^*$$

$$= \sum f(n_i; p_i) \int_0^{A(n_i)} g(s^*/n_i) ds^*$$

$$= \sum_{n_i} f(n_i;\ p_i) \cdot \alpha = \alpha \cdot \sum_{n_i} f(n_i;\ p_i) = \alpha$$

where s^* represents λ or W and $\sum_{n_i} f(n_i;\ p_i) = 1$. Hence, to make a test one needs only to find an $A(n_1, \ldots, n_t)$, such that

$$\int_0^{A(n_1, \ldots, n_t)} g(s^*/n_1, \ldots, n_t)\, ds^* = \alpha$$

But, $g(s^*/n_1, \ldots, n_t)$ is just a beta distribution in terms of W or a function of an F variate in terms of λ . This, then, implies that to test the hypothesis $H_o(\mu_1 = \mu_2 = \ldots = \mu_t = \mu_0)$ when the sample variances are the same and when the sample sizes are assumed to be random variables, one uses the conventional F - test method. In other words the likelihood-ratio will have an F distribution.

To fully examine this test, however, one should examine the power function $g(\theta)$. But, in order to be able to do this, it is necessary to actually be able to evaluate (18) . At the moment, this is not possible, since the expression is too complex, or seems to be. Until an approximate expression can be found or this expression simplified to a point that (18) can be evaluated, the power function will not be able to be evaluated. This could develop into a thesis in and of itself.

CONCLUSIONS

In summary, a density has been found which describes the experiment when the sample sizes are assumed to be random variables. The likelihood-ratio was used to test the hypothesis $H_o(\mu_1 = \mu_2 = \ldots = \mu_t = \mu_0)$.

A mathematical relationship was obtained for the density of the likelihood-ratio criterion, λ , which was too complex to obtain the necessary probabilities for testing λ . It was shown, however, that the conventional F test could be used to make a test for λ , even though it is assumed that the sample sizes are random variables.

The power function, β , was not compared with the known power functions for fixed sample sizes due to the complexity of the aforementioned density function.

LITERATURE CITED

Cochran, W. G. and G. M. Cox. 1957. Experimental designs. 2nd Ed. John Wiley and Sons, New York.

Cox, D. R. 1958. Planning of Experiments. John Wiley and Sons, New York.

Davies, O. L. (ed). 1954. The design and analysis of industrial experiments. Oliver and Boyd, London.

Federer, W. T. 1955. Experimental design. Macmillan Co., New York.

Feller, William. 1957. An introduction to probability theory and its applications. Volume I. John Wiley and Sons, Inc., New York, London.

Finney, D. J. 1960. An introduction to the theory of experimental design. The University of Chicago Press, Chicago, Illinois.

Fisher, R. A. 1925. Statistical methods for research workers. Oliver and Boyd, Ltd., London and Edinburgh.

Graybill, Franklin A. 1961. An introduction to linear statistical models. Volume I. McGraw-Hill Book Company, Inc., New York, Toronto, London.

Kempthorne, O. 1952. The design and analysis of experiments. John Wiley and Sons, New York.

Kendal, M. G. 1951. The advanced theory of statistics. Volume II. Hafner Publishing Company, New York.

Mood, Alexander M. and Franklin A. Graybill. 1963. Introduction to the theory of statistics. McGraw-Hill Book Company, Inc., New York, San Francisco, Toronto, London.

Parzen, Emanuel. 1960. Modern probability theory and its applications. John Wiley and Sons, New York, London.

Quenouille, M. H. 1953. The design and analysis of experiments. Charles Griffin and Company, London.