A GENERALIZATION OF ANALYSIS OF VARIANCE AND MULTIVARIATE ANALYSIS TO DATA BASED ON FREQUENCIES IN QUALITATIVE CATEGORIES OR CLASS INTERVALS

by

S. N. Roy and Marvin A. Kastenbaum

This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command, and in part by the United States Public Health Service HTS 5065 (c).

Institute of Statistics
Mimeograph Series No. 131
June 1, 1955

UNCLASSIFIED

Security Information

Bibliographical Control Sheet

1. O.A.: Institute of Statistics, North Carolina State College of the University of North Carolina

   M.A.: Office of Scientific Research of the Air Research and Development Command

2. O.A.: CIT Report No. 19

   M.A.: OSR-TN-55-167

3. A GENERALIZATION OF ANALYSIS OF VARIANCE AND MULTIVARIATE ANALYSIS TO DATA BASED ON FREQUENCIES IN QUALITATIVE CATEGORIES OR CLASS INTERVALS

4. S. N. Roy and Marvin A. Kastenbaum

5. June, 1955

6. 27

7. None

8. AF 18(600)-83

9. RDO No. 670-088

10. UNCLASSIFIED

11. None

12. In the situation indicated by the title a p-variate body of data arranged in a q-way classification will formally look like a body of data arranged in a (p + q)-way table, but the fundamental distinction between a so-called "variate" and a so-called "way of classification" is that along the direction of a "variate," the marginal frequencies are fixed or prescribed. The hypotheses of "no total correlation," "no multiple correlation," "no partial correlation," "no canonical correlation," "no main effect," "no interaction," etc., are translated into hypotheses on the structure of the probabilities over the different cells or categories, and, with large sample assumptions, these hypotheses are tested by χ^2 with appropriate degrees of freedom. No exact test in terms of the original multinomial distribution is attempted in this paper.

A GENERALIZATION OF ANALYSIS OF VARIANCE AND MULTIVARIATE ANALYSIS TO DATA BASED ON FREQUENCIES IN QUALITATIVE CATEGORIES OR CLASS INTERVALS[1]

by

S. N. Roy and Marvin A. Kastenbaum[2]

1. Summary. In the situation indicated by the title a p-variate body of data arranged in a q-way classification will formally look like a body of data arranged in a (p + q)-way table, but the fundamental distinction between a so-called "variate" and a so-called "way of classification" is that along the direction of a "variate," the marginal frequencies are stochastic variates, while along a "way of classification," the marginal frequencies are fixed or prescribed. The hypotheses of "no total correlation," "no multiple correlation," "no partial correlation," "no canonical correlation," "no main effect," "no interaction," etc., are translated into hypotheses on the structure of the probabilities over the different cells or categories, and, with large sample assumptions, these hypotheses are tested by χ^2 with appropriate degrees of freedom. No exact test in terms of the original multinomial distribution is attempted in this paper.

2. Notation and Preliminaries. To fix our ideas, consider a [redacted] sample of size n, distributed over a three-way table in terms of, let us assume for the moment, three variates. Let n_{ijk} denote the observed frequency, and p_{ijk}, the probability under any given hypothesis of having an observation in the (ijk)-th cell, where $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$. Also let the marginals

---

1. This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command, and in part by the United States Public Health Service HTS 5065 (c).

2. Marvin A. Kastenbaum: Research Fellow in the Department of Biostatistics, University of North Carolina

be denoted by $\sum_i n_{ijk} = n_{ojk}$, $\sum_j n_{ijk} = n_{iok}$, $\sum_k n_{ijk} = n_{ijo}$, $\sum_{i,j} n_{ijk} = n_{ook}$, $\sum_{i,k} n_{ijk} = n_{ojo}$, $\sum_{j,k} n_{ijk} = n_{ioo}$, $\sum_{i,j,k} n_{ijk} = n$(say). Let the corresponding summations over p_{ijk} be denoted by p_{ojk}, p_{iok}, p_{ijo}, p_{ook}, p_{ojo}, p_{ioo}, p_{ooo}. Since the categories are mutually exclusive, it is easy to see that these are, in fact, the corresponding marginal probabilities, so that $p_{ooo} = 1$. The generalization to more than three variates would be obvious. The total number n is, in any case, supposed to be fixed. The likelihood function, which in this case is also the probability of the n_{ijk}'s, is given by

$$(2.1) \qquad \phi(n_{ijk}\text{'s}) = \phi \text{ (say)} = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{\,n_{ijk}} \sim \prod_{i,j,k} p_{ijk}^{\,n_{ijk}} .$$

The last expression on the right side of (2.1) is the one we shall need when we are interested in finding the maximum likelihood estimates of the p's.

We have $E(n_{ijk}) = np_{ijk}$, from which it is easy to see, by taking summation, that $E(n_{ijo}) = np_{ijo}$, $E(n_{ioo}) = np_{ioo}$, etc., and, in general, for any linear function of n's the same type of relationship will hold for the p's. Starting from (2.1), we next observe that the conditional probabilities of the n_{ijk}'s, given, say n_{ioo}'s ($i = 1, 2, \ldots, r$), or say n_{ijo}'s ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$), will be given respectively by

$$(2.2) \qquad \phi(n_{ijk}\text{'s} \mid n_{ioo}\text{'s}) = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{\,n_{ijk}} \Bigg/ \frac{n!}{\prod_{i} n_{ioo}!} \prod_{i} p_{ioo}^{\,n_{ioo}} ,$$

and

$$(2.3) \qquad \phi(n_{ijk}\text{'s} \mid n_{ijo}\text{'s}) = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{\,n_{ijk}} \Bigg/ \frac{n!}{\prod_{i,j} n_{ijo}!} \prod_{i,j} p_{ijo}^{\,n_{ijo}} .$$

If now the n_{ioo}'s are held fixed, or say the n_{ijo}'s are held fixed, we shall have a self consistent set up if we put $p_{ioo} = n_{ioo} / n$, or, in the second case, $p_{ijo} = n_{ijo} / n$, and also take the right sides of (2.2) and (2.3) to be the actual probabilities of the n_{ijk}'s in the two different situations. The generalization to more general types of linear constraints on the n's is obvious. Notice that if, for example, the n_{ijo}'s are held fixed, and if we want to estimate the p's by, say, the maximum likelihood method, we do not have to estimate the p_{ijo}'s since they are already given by $p_{ijo} = n_{ijo}/n$. For this purpose, therefore, it is enough to replace (2.3) by

$$(2.4) \qquad \phi \sim \prod_{i,j,k} p_{ijk}^{n_{ijk}} ,$$

and obtain the maximum likelihood estimates of the p_{ijk}'s subject to $\sum_k p_{ijk}$ being fixed (with i = 1, 2, ..., r; j = 1, 2, ..., s). Likewise when the n_{ioo}'s are held fixed, it is enough to replace (2.2) by (2.4) and obtain the maximum likelihood estimates of the p_{ijk}'s subject to $\sum_{j,k} p_{ijk}$ being fixed (with i = 1, 2, ..., r). The generalization of this to other linear constraints on the n's is also obvious. Notice that any linear constraint on the n's will imply a similar linear constraint on the p's, but it is not the other way around.

3. Hypotheses of Independence between "i" and "j" in a Two-Way Table. [2,4,5]

$$(3.1) \qquad H_0: \; p_{ij} = p_{io}\, p_{oj} \quad \text{against } H: \; p_{ij} \neq p_{io}\, p_{oj} \quad (\text{for } i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, s).$$

Case I: Both "i" and "j" are "variates"; or in other words, the hypothesis is a composite one, where the p_{io}'s and p_{oj}'s are the free or nuisance parameters, subject to $\sum_i p_{io} = \sum_j p_{oj} = 1$. We shall see that in this situation (r + s - 2) independent and free parameters have to be eventually estimated from the observations. There is also one linear constraint on the n's which is $\sum_{i,j} n_{ij} = n$. (3.1) is the analogue of "no correlation" for a bivariate normal population.

Case II: Either "i" or "j", say "i", is a "way of classification", while "j" is a "variate". In other words the n_{io}'s are fixed but the n_{oj}'s are stochastic variates. In this case $p_{io} = n_{io}/n$, but, of course $p_{oj} \neq n_{oj}/n$. Thus the only free and nuisance parameters are the p_{oj}'s subject to $\sum_j p_{oj} = 1$, so that (s - 1) free parameters have to be eventually estimated from the data, and we have n_{io}'s (i = 1, 2, ..., r) all fixed; that is, we have r linear constraints on the n_{ij}'s. Physically, H_0 of (3.1) in this case means testing the hypothesis that r observed frequency distributions with fixed marginal totals, n_{io}'s, have come from the same parent frequency distribution. This is easily seen to be one natural generalization of the hypothesis of the equality of means for a one-way classification in the analysis of variance, when we remember that $H_0: \xi_1 = \xi_2 = \ldots = \xi_r$ for $N(\xi_i, \sigma^2)$ (i = 1, 2, ..., r) would imply that the r distributions are the same. For the normal distribution the class of alternatives is supposed to be $H \neq H_0$ under the model $N(\xi_i, \sigma^2)$, but here the class of alternatives is, of course, much more general. The case "i" being a "variate" and "j" a "way of classification" is exactly similar, and need not be considered separately.

Case III: Both "i" and "j" are "ways of classification". Here the sets of n_{io}'s and n_{oj}'s are both fixed so that there are (r - 1) + (s - 1) + 1 independent linear constraints on the n_{ij}'s, while $p_{io} = n_{io}/n$ and $p_{oj} = n_{oj}/n$, so that no parameter has to be estimated, all of them being prescribed. This is the case usually given in the textbooks, and this is exactly the case which is most difficult to visualize, unless we think in terms of a hypothetical sub-population having the same fixed marginals as the ones we have observed, which is anyway a highly artificial concept. However, for an r x r case, an extension of Fisher's "tea-tasting" experiment would provide a realistic example of fixed marginals both ways. But, it will be shown later, that, in all three cases we end up with the same test of H_0 of (3.1). This is a highly interesting result.

4. Hypotheses Associated with a Three-Way Table: [6] Features of the three-way table which, by considering the marginals, can be easily seen to be identical with those of a two-way classification, are not of so much interest. We shall, therefore, restrict ourselves mainly to those hypotheses which have no analogue in a two-way table. Also, out of the possible cases (I) "i", "j", and "k" all being "variates"; (II) any two of them being "ways of classification" and the remaining one a "variate"; (III) any one of them being a "way of classification", and the remaining two "variates"; (IV) all of them being "ways of classification", we shall discuss, in the present paper, (I) and (II), these being of greater physical interest than the others. Mathematically, however, it will turn out that each hypothesis will have the same test for all the different cases.

4.1 Hypotheses of Conditional Independence between "i and j" | "k". It is easy to see that the conditional probability of "i and j" | "k" = p_{ijk}/p_{ook}, and the conditional probabilities of "i" | "k" and "j" | "k" are respectively p_{iok}/p_{ook} and p_{ojk}/p_{ook}. Thus the hypothesis of conditional independence between "i and j" | "k" is

$$(4.1.1) \qquad H_0: \frac{p_{ijk}}{p_{ook}} = \frac{p_{iok}}{p_{ook}} \cdot \frac{p_{ojk}}{p_{ook}} \quad \text{or} \quad p_{ijk} = \frac{p_{iok}p_{ojk}}{p_{ook}},$$

(for i = 1, 2, ..., r; j = 1, 2, ..., s; k = 1, 2, ..., t). The alternative class is, of course, $H \neq H_0$. In (4.1.1) adding up over i and j respectively we have

$$(4.1.2) \qquad p_{ojk} = p_{ojk} \qquad \text{and} \quad p_{iok} = p_{iok},$$

which are merely two consistency conditions. Adding up over k, we have

$$(4.1.3) \qquad p_{ijo} = \sum_k \frac{p_{iok}p_{ojk}}{p_{ook}} .$$

If on (4.1.1) we superimpose the conditions of independence between "i" and "k" and "j" and "k", i.e., (4.1.4) $$p_{iok} = p_{ioo}p_{ook} \quad \text{and} \quad p_{ojk} = p_{ojo}p_{ook},$$

we have

(4.1.5) $$p_{ijk} = p_{ioo}p_{ojo}p_{ook},$$

which is the condition of complete independence of "i", "j", and "k". Notice that (4.1.1) will not imply (4.1.4), and (4.1.4) by itself will not imply (4.1.5).

Now consider for (4.1.1) the Case I, where "i", "j", and "k" are all variates. The H_0 of (4.1.1) is now easily seen to be the analogue of "no partial correlation" between the first two variates, in the case of a three-variate normal population. Now if we want to test (4.1.1) we must eventually estimate the nuisance parameters p_{iok}'s, p_{ojk}'s, and p_{ook}'s subject to $\sum_i p_{iok} = \sum_j p_{ojk} = p_{ook}$ $(k = 1, 2, \ldots, t)$, and $\sum_k p_{ook} = 1$. It is easy to check that this leaves us with rt + st + t - t - t - 1, that is t(r + s - 1) - 1 free parameters to estimate. We have just one linear constraint on the n's, namely $\sum_{i,j,k} n_{ijk} = n$ (fixed).

If we start out to test (4.1.5), we would have to estimate, eventually, the nuisance parameters p_{ioo}'s, p_{ojo}'s, and p_{ook}'s subject to $\sum_i p_{ioo} = \sum_j p_{ojo} = \sum_k p_{ook} = 1$, i.e., (r + s + t - 3) independent parameters. There is also the same one linear constraint on the n's, as stated above in the case of (4.1.1).

It will be seen in sections 5, 7, and 8 that it is unrealistic to try to test (4.1.1) for the Case II, where "i" and "j" are "ways of classification" and "k" is a "variate".

4.2 <u>Hypothesis of Independence between "(i,j)" and "k"</u>, that is, the hypothesis of multiple independence. Consider

(4.2.1) $$H_0\colon\ p_{ijk} = p_{ijo}p_{ook} \qquad (\text{for } i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, s;\ k = 1, 2, \ldots, t),$$

the alternative being, of course, $H \neq H_0$. It is easy to check, by summing over i and j respectively that (4.2.1) implies

(4.2.2) $$p_{iok} = p_{ioo}p_{ook} \quad \text{and} \quad p_{ojk} = p_{ojo}p_{ook}.$$

Summing over k we have merely the consistency condition

(4.2.2.1) $$p_{ijo} = p_{ijo}.$$

Notice that although (4.2.1) implies (4.2.2), the condition (4.2.2) will not, in general, imply (4.2.1). However, for a normal population (4.2.2) implies (4.2.1). Let us ask ourselves what set of conditions is there which, when superimposed on (4.2.2) will, together, be exactly equivalent to (4.2.1). One possible set might appear to be

(4.2.3) $$H_0: \; p_{ijk} = \frac{p_{ijo}p_{iok}p_{ojk}}{p_{ioo}p_{ojo}p_{ook}} \quad \text{(for } i = 1, 2, \ldots, r; \; j = 1, 2, \ldots, k = 1, 2, \ldots, t).$$

Check that (4.2.3) does not imply (4.2.2), but that if on (4.2.3) we superimpose (4.2.2), we have (4.2.1) all right. But (4.2.3) would be mathematically most difficult to handle, in that the parameters on the right side of this equation are subject to sets of side conditions, typical among them being

(4.2.4) $$\sum_k p_{ijk} = p_{ijo} = \sum_k \frac{p_{ijo}p_{iok}p_{ojk}}{p_{ioo}p_{ojo}p_{ook}}$$

or

$$\sum_k \frac{p_{iok}p_{ojk}}{p_{ook}} = p_{ioo}p_{ojo},$$

and other such sets obtained by permuting the subscripts. In fact, (4.2.3) was tried, and was found to be intractable.

Physically a less natural and more abstract, but mathematically a much easier set of conditions seems to be

(4.2.5) $$H_0: p_{ijk} = \frac{q_{ijo}q_{iok}q_{ojk}}{q_{ioo}q_{ojo}q_{ook}} \quad (i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, s;\ k = 1, 2, \ldots, t),$$

where we do not assume that $q_{ijo} = p_{ijo}$ etc., nor even that $q_{ioo} = \sum_j q_{ijo}$, etc. Equation (4.2.5), after elimination of the q's, leads to a number of constraints on the p's themselves, and it is easier to try to estimate the p's subject to these constraints and to $\sum_{i,j,k} p_{ijk} = 1$, rather than to try to estimate the q's. The only role of the q's and of the hypothesis (4.2.5) is one of yielding certain constraints on the p's themselves. It will be shown in sections 8 and 9 that (4.2.5) is equivalent to just (r - 1) (s - 1) (t - 1) constraints on the p_{ijk}'s, which, together with $\sum_{i,j,k} p_{ijk} = 1$, make just (r - 1) (s - 1) (t - 1) + 1 constraints. Notice that in this case we do not have constraints like (4.2.4) which, in practice, turn out to be quite awkward. Now we lay down the rule, which is physically rather abstract, but mathematically quite straightforward, that if (4.2.2) is true, that is, if the hypothesis (4.2.2) is tested and accepted, we shall substitute in (4.2.5) p_{ijo}, p_{iok}, etc. for q_{ijo}, q_{iok}, etc., and p_{ioo} for q_{ioo} and so on, and superimpose (4.2.2), and end up with (4.2.1). Notice that if in (4.2.5) we were to replace (i,j,k) by (x,y,z), then (4.2.5) would be found to imply

(4.2.6) $$f(x,y,z) = \frac{f_1(x,y)\ f_2(x,z)\ f_3(y,z)}{F_1(x)\ F_2(y)\ F_3(z)}$$

with nothing else connecting f_1, f_2, f_3, F_1, F_2, F_3 among themselves or with f.

We shall now consider cases I and II, each in relation to (4.2.1) and (4.2.5).

4.3 Case I. "i", "j", and "k" are all "variates": In this case, (4.2.1) is the natural analogue of the hypothesis of "no multiple correlation" between "(i,j)" and "k". We have to estimate the nuisance parameters p_{ijo}'s and p_{ook}'s subject to $\sum_{i,j} p_{ijo} = \sum_k p_{ook} = 1$, which leaves us with (rs - 1) + (t - 1) free parameters to be

estimated. There is, of course, the linear constraint on the n's, namely
$\sum_{i,j,k} n_{ijk} = n$ (fixed).

Turning now to (4.2.5) as applied to Case I we notice that this does not have any analogue in multivariate analysis based on the normal population. We shall find, in the next subsection on Case II, where "i" and "j" are "ways of classification" and "k" is a "variate", that (4.2.5) is really the hypothesis of "no interaction". For Case I, we can thus regard (4.2.5) as a contribution to multivariate analysis made by analysis of variance. From the remarks on the constraints on the p_{ijk}'s following from (4.2.5), we note that the number of free parameters to be estimated is $rst - [(r-1)(s-1)(t-1)+1]$, and we have the usual linear constraint on the n's, namely $\sum_{i,j,k} n_{ijk} = n$ (fixed).

4.4 <u>Case II</u>. <u>"i" and "j" are "ways of classification" and "k" is a "variate"</u>. In this case the n_{ijo}'s are fixed by the conditions of the experiment, which means that $p_{ijo} = \frac{n_{ijo}}{n}$, and thus the p_{ijo}'s do not have to be estimated from the data. (4.2.1), in this case, will thus be the hypothesis of equal (population) frequency distributions, in terms of the "variate" "k" over the r x s categories. To test (4.2.1), there are p_{ook}'s to be estimated subject to $\sum_k p_{ook} = 1$, which means that we have $t - 1$ free parameters to estimate. We have also r x s linear constraints on the n's, by virtue of the n_{ijo}'s being given.

Turning now to (4.2.5) as applied to Case II, we note that if

(4.4.1) $$H_0: \; p_{iok} = p_{ioo}p_{ook} \quad \text{and} \quad H_0: \; p_{ojk} = p_{ojo}p_{ook}$$

are tested and accepted, that is, if in the (ij) classification, the marginal i's (i = 1, 2, ..., r) have the same frequency distribution in terms of "k", and so also the marginal j's (j = 1, 2, ..., s), then substituting p_{ijo}'s, etc. for q_{ijo}'s etc. in (4.2.5) and then superimposing (4.4.1) on (4.2.5) we have

(4.4.2) $$p_{ijk} = p_{ijo}p_{ook} \quad \text{(over all } i, j, k).$$

This means that in every (ij)-cell there is the same frequency distribution in terms of "k".

Going back to the usual model of analysis of variance for a two way classification, and assuming, for simplicity, equal frequencies, say u, over the different (ij) cells, we recall that if $x_{ij\lambda}$ be the λ-th observation in the (ij)-th cell (λ = 1, 2, ..., u), we assume that $x_{ij\lambda}$ is $N[E(x_{ij\lambda}), \sigma^2]$, and

(4.4.3) $$E(x_{ij\lambda}) = \mu_{oo} + \mu_{io} + \mu_{oj} + \mu_{ij},$$

where $$\sum_i \mu_{io} = \sum_j \mu_{oj} = \sum_i \mu_{ij} = \sum_j \mu_{ij} = 0.$$

The condition of "no interaction" is usually expressed as

(4.4.4) $$\mu_{ij} = 0 \qquad (i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, s),$$

and the condition for "no main 'i'-effect" as

(4.4.5) $$\mu_{io} = 0 \qquad (i = 1, 2, \ldots, r),$$

and that for "no main 'j'-effect" as

(4.4.6) $$\mu_{oj} = 0 \qquad (j = 1, 2, \ldots, s).$$

If all these hold, then we have

(4.4.7) $$x_{ij\lambda}:\ N(\mu_{oo}, \sigma^2),$$

for all i, j, and λ, which means that every (ij) cell has the same frequency distribution in terms of the variate x, the distribution in this case being normal. It is easy to see that (4.4.2) is a proper generalization of (4.4.7). It is now easy to check that none of the three conditions (4.4.4), (4.4.5), and (4.4.6) implies the other two, and no two of these imply the third, and also that none or no two of them, separately, will imply (4.4.7). All of these have to be true in order to lead to (4.4.7). Assuming now any "k" interval to be the interval between x and x + dx, and remembering that $p_{ijo} = u/n = u/urs = 1/rs$, we have in general

(4.4.8) $$p_{ijk}/p_{ijo} = (1/\sigma\sqrt{2\pi}) \exp [-1/2\sigma^2 \{x - E(x)\}^2] dx,$$

or

$$p_{ijk} = (1/rs\sigma\sqrt{2\pi}) \exp [-1/2\sigma^2 \{x - E(x)\}^2] dx,$$

where E(x) is given by the right side of (4.4.3). If now (4.4.4) and (4.4.5) hold, but not necessarily (4.4.6), then substituting from (4.4.4) and (4.4.5) in (4.4.3), and summing the two sides of (4.4.8) over j, we have

(4.4.9) $$p_{iok} = (1/rs\sigma\sqrt{2\pi}) \sum_j \exp [-1/2\sigma^2 \{x - \mu_{oo} - \mu_{oj}\}^2] dx.$$

Notice that this is independent of "i", which means that in every "i"-cell there is a distribution in terms of "k" or x which is the same for all "i". It is obvious that there is a similar result for "j" after summing over "i". Notice now that (4.4.1) is a generalization of these. Thus we can regard (4.4.1) as one analogue of "no main effects" and (4.2.5) as one analogue of "no interaction". The reader will easily perceive that this generalization does not retain all the detailed features of analysis of variance as we know it in the case of the highly structured normal populations. But we believe that some important features are retained.

5. Large Sample Tests in terms of χ^2. [4, 5] It is well known that if (i) there is a total of n observations distributed over s cells such that the number of observations in the j-th cell is n_j ($j = 1, 2, \ldots, s$), and if (ii) the n_j's are subject to the linear constraints

(5.1) $$\sum_{j=1}^{s} \lambda_{ij} n_j = \mu_i \qquad (i = 1, 2, \ldots, r < s),$$

where $\lambda(r \times s)$ is of rank r, and if (iii) the probability p_j $(j = 1, 2, \ldots, s)$ of an observation in the j-th cell be of the form $p_j(\theta_1, \theta_2, \ldots, \theta_t)$, where $r + t < s$, and if (iv) the θ_k's are estimated by the modified minimum χ^2 method (which has been shown to be the same as that of maximum likelihood), and if (v) this leads to a unique solution in the p_j's, to be called, say $\hat{p}_j$'s, then for reasonably large n_j's,

$$(5.2) \qquad \sum_{j=1}^{s} (n_j - n\hat{p}_j)^2/n\hat{p}_j$$

is approximately distributed as a χ^2 with degrees of freedom equal to $s - r - t$. That is to say, the degrees of freedom of χ^2 are equal to the number of cells minus the number of independent linear constraints on the n_j's, minus the number of free parameters to be estimated from the data. Notice that (5.1) includes the condition $\sum_j n_j = n$. Notice also that every linear constraint on (5.1) will imply a similar constraint on the p_j's (although the reverse is not true), and will thus reduce the number of free parameters θ_k's. Notice further that customarily the role of any hypothesis (that we test by χ^2) is to give a structure of p_j's in terms of θ_k's, which then have to be estimated from the data.

6. Applications of the χ^2-test to Section 3. Starting from (3.1) and taking into account the remarks of section 2, we write the likelihood function as

$$(6.1) \qquad \phi \sim \prod_{i,j} p_{ij}^{n_{ij}}, \text{ that is, } \phi \sim \prod_i p_{io}^{n_{io}} \prod_j p_{oj}^{n_{oj}} .$$

Now let us consider the three cases separately.

Case I. We must estimate both the p_{io}'s and the p_{oj}'s, subject to $\sum_i p_{io} = \sum_j p_{oj} = 1$. Introducing the usual Lagrangian multipliers λ and μ, and taking the logarithm of the right side of (6.1), we have as the maximum likelihood equations for the p_{io}'s and the p_{oj}'s

$$(6.2) \qquad \frac{n_{io}}{p_{io}} + \lambda = 0 \quad (i = 1, 2, \ldots, r), \text{ and}$$

$$(6.3) \qquad \frac{n_{oj}}{p_{oj}} + \mu = 0 \quad (j = 1, 2, \ldots, s).$$

Multiplying both sides of (6.2) by p_{io} and summing over i, and using the side condition that $\sum_i p_{io} = 1$, and also multiplying both sides of (6.3) by p_{oj} and summing over j, and using the side condition $\sum_j p_{oj} = 1$, it is immediately seen that

$$(6.4) \qquad \lambda = \mu = -n;$$

so that we have the maximum likelihood estimates of p_{io} and p_{oj} given by

$$(6.5) \qquad \hat{p}_{io} = n_{io}/n \quad \text{and} \quad \hat{p}_{oj} = n_{oj}/n.$$

Substituting in the usual expression for χ^2, we have the modified χ^2 given by

$$(6.6) \qquad \sum_{i,j} (n_{ij} - n\hat{p}_{io}\hat{p}_{oj})^2 / n\hat{p}_{io}\hat{p}_{oj} \; .$$

Recalling section 5 and the observations under case I of section 3, we note that (6.6) has a χ^2-distribution with d.f. $rs - 1 - (r + s - 2) = (r - 1)(s - 1)$.

Case II. Here $p_{io} = n_{io}/n$, and the p_{oj}'s alone have to be estimated under the side condition $\sum_j p_{oj} = 1$. Proceeding as in the previous case, we observe that here we shall have only the equation (6.3), and we end up with

$$(6.7) \qquad p_{io} = n_{io}/n \quad \text{and} \quad \hat{p}_{oj} = n_{oj}/n.$$

Notice the difference between (6.5) and (6.7). In (6.5) there are carats on both p_{io} and p_{oj}, while in (6.7) the carat appears only on p_{oj}. Now substituting in the usual expression for χ^2, we get

(6.9) $$\sum_{i,j} (n_{ij} - np_{io}\hat{p}_{oj})^2 / np_{io}\hat{p}_{oj} \ .$$

Recalling the observations under Case II of section 3, we have a χ^2 with d.f. $rs - r - (s - 1) = (r - 1)(s - 1)$.

Case III. Here we have

(6.10) $$p_{io} = n_{io}/n \quad \text{and} \quad p_{oj} = n_{oj}/n.$$

Substituting in the usual expression for χ^2, we get

(6.11) $$\sum_{i,j} (n_{ij} - np_{io}p_{oj})^2 / np_{io}p_{oj} \ ,$$

and recalling the observations made under case III of section 3, we have χ^2 with d.f. $rs - (r + s - 1) = (r - 1)(s - 1)$.

The familiar text book example of "vaccinated" against "not vaccinated" one way, and "attacked" against "not attacked" the other way is really an example of Case II; it is neither Case I nor Case III.

7. Application of the χ^2-test to section 4. In this section we shall use the χ^2 to test all the different hypotheses in section 4 except the hypothesis (4.2.5) which, as we have already observed in section 4, plays the role of "no interaction" when "i" and "j" are "ways of classification" and "k" is a "variate", and a certain role, which has no analogue in multivariate analysis when all three are "variates". The hypothesis (4.2.5) will be considered in detail in section 8 for the special case of $r = s = t = 2$, and in section 9, in less detail, for the general $r \times s \times t$ table.

Conditional Independence. Starting from (4.1.1), and taking into account the remarks in section 2, we can write down the likelihood function as

(7.1) $$\phi \sim \prod_{i,j,k} p_{ijk}^{\ n_{ijk}}, \text{ that is, } \phi \sim \prod_{i,k} p_{iok}^{\ n_{iok}} \prod_{j,k} p_{ojk}^{\ n_{ojk}} \prod_{k} p_{ook}^{\ -n_{ook}} \ .$$

<u>Case I</u>. <u>"i", "j", and "k" all are "variates"</u>. Notice that we have to estimate the p_{iok}'s, p_{ojk}'s, and p_{ook}'s subject to

$$(7.2) \qquad \sum_i p_{iok} = \sum_j p_{ojk} = p_{ook} \quad (k = 1, 2, \ldots, t) \text{ and } \sum_k p_{ook} = 1.$$

Now using the same method as in section 6, and calling the associated Lagrangian multipliers λ_k, μ_k $(k = 1, 2, \ldots, t)$, and ν, we have the maximum likelihood equations

$$(7.3) \qquad \frac{n_{iok}}{p_{iok}} + \lambda_k = 0; \quad \frac{n_{ojk}}{p_{ojk}} + \mu_k = 0; \quad \frac{n_{ook}}{p_{ook}} + \lambda_k + \mu_k - \nu = 0,$$

with $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$. Multiply the first equation of (7.3) by p_{iok} and sum over i using the side condition $\sum_i p_{iok} = p_{ook}$: multiply the second equation of (7.3) by p_{ojk} and sum over j using the side condition $\sum_j p_{ojk} = p_{ook}$. Now using the third equation of (7.3) we have

$$(7.4) \qquad \lambda_k = \mu_k = -n_{ook} / p_{ook} = \nu.$$

Multiplying the third equation by p_{ook} and summing over k using the side condition $\sum_k p_{ook} = 1$, we have

$$(7.5) \qquad \nu = -n.$$

Hence we have the following maximum likelihood estimates:

$$(7.6) \qquad \hat{p}_{iok} = \frac{n_{iok}}{n} ; \quad \hat{p}_{ojk} = \frac{n_{ojk}}{n} ; \quad \hat{p}_{ook} = \frac{n_{ook}}{n} .$$

Substituting in the usual expression for X^2 we get

$$(7.7) \qquad \sum_{i,j,k} \left(n_{ijk} - n \frac{\hat{p}_{iok}\hat{p}_{ojk}}{\hat{p}_{ook}}\right)^2 / n \frac{\hat{p}_{iok}\hat{p}_{ojk}}{\hat{p}_{ook}} .$$

Recalling the observations under Case I of sub-section 4.1, we have a χ^2 with d.f. $rst - 1 - [t(r + s - 1) - 1] = t(r - 1)(s - 1)$.

The question as to how far it is meaningful to investigate this "conditional independence" for the other cases, namely, when not all are "variates", is now under examination. However, there are additional mathematical difficulties in these situations.

Complete Independence. Starting from (4.1.5) and recalling section 2, we write the likelihood function as

$$(7.8) \qquad \phi \sim \prod_{i,j,k} p_{ijk}^{n_{ijk}}, \text{ that is, } \phi \sim \prod_i p_{ioo}^{n_{ioo}} \prod_j p_{ojo}^{n_{ojo}} \prod_k p_{ook}^{n_{ook}}.$$

Case I. "i", "j", and "k" all are variates. Here we have to estimate the p_{ioo}'s, p_{ojo}'s, and p_{ook}'s under the side conditions $\sum_i p_{ioo} = \sum_j p_{ojo} = \sum_k p_{ook} = 1$. Introducing the usual Lagrangian multipliers, we have as our maximum likelihood equations

$$(7.9) \qquad \frac{n_{ioo}}{p_{ioo}} + \lambda = \frac{n_{ojo}}{p_{ojo}} + \mu = \frac{n_{ook}}{p_{ook}} + \nu = 0,$$

with $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$. Solving we have

$$(7.10) \qquad \lambda = \mu = \nu = -n, \quad \text{and thus}$$

$$(7.11) \qquad \hat{p}_{ioo} = \frac{n_{ioo}}{n}; \quad \hat{p}_{ojo} = \frac{n_{ojo}}{n}; \quad \hat{p}_{ook} = \frac{n_{ook}}{n}.$$

Substituting this in the usual χ^2 expression we have

$$(7.12) \qquad \sum_{i,j,k} (n_{ijk} - n\hat{p}_{ioo}\hat{p}_{ojo}\hat{p}_{ook})^2 / n\hat{p}_{ioo}\hat{p}_{ojo}\hat{p}_{ook},$$

and recalling the observations under Case I of sub-section (4.1) we see that our χ^2 has d.f. $rst - 1 - (r + s + t - 3) = rst - r - s - t + 2$.

In this case, as in section 6, it can be shown that we should have the same χ^2 with the same d.f. for (i) any two of "i", "j", and "k" as "variates" - say "i" and "j" - and "k" as a "way of classification", (ii) any one of "i", "j", and "k", say "i", as a "variate" and "j" and "k" as "ways of classification", or (iii) all three as "ways of classification". In case (i) we should have $\hat{p}_{ioo} = n_{ioo}/n$; $\hat{p}_{ojo} = n_{ojo}/n$; and $p_{ook} = n_{ook}/n$; in case (ii) $\hat{p}_{ioo} = n_{ioo}/n$; $p_{ojo} = n_{ojo}/n$; and $p_{ook} = n_{ook}/n$; and in case (iii) $p_{ioo} = n_{ioo}/n$; $p_{ojo} = n_{ojo}/n$; and $p_{ook} = n_{ook}/n$.

<u>Multiple independence between "(i, j)" and "k"</u>. Starting from (4.2.1) we write down the likelihood function as

$$(7.13) \qquad \phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j} p_{ijo}^{\,n_{ijo}} \prod_{k} p_{ook}^{\,n_{ook}} \sim \prod_{i,j} p_{ijo}^{\,n_{ijo}} \prod_{k} p_{ook}^{\,n_{ook}}.$$

<u>Case I</u>. <u>"i", "j", and "k" all are "variates"</u>. We have to estimate the p_{ijo}'s and p_{ook}'s subject to the constraints $\sum_{i,j} p_{ijo} = \sum_{k} p_{ook} = 1$. Introducing the usual Lagrangian multipliers, we have as our maximum likelihood equations

$$(7.14) \qquad \frac{n_{ijo}}{p_{ijo}} + \lambda = 0 \quad \text{and} \quad \frac{n_{ook}}{p_{ook}} + \mu = 0$$

for (i = 1, 2, ..., r; j = 1, 2, ..., s) and (k = 1, 2, ..., t). Solving, we have

$$(7.15) \qquad \hat{p}_{ijo} = n_{ijo}/n \quad \text{and} \quad \hat{p}_{ook} = n_{ook}/n,$$

and substituting in the usual χ^2 expression we get

$$(7.16) \qquad \sum_{i,j,k} (n_{ijk} - n\hat{p}_{ijo}\hat{p}_{ook})^2 / n\hat{p}_{ijo}\hat{p}_{ook},$$

which, recalling the remarks under Case I of subsection (4.3) is a χ^2 with d.f. $rst - 1 - [(rs - 1) + (t - 1)] = (rs - 1)(t - 1)$. We shall end up with the same χ^2 with the same d.f. in the cases where (i) "i" and "j" are "variates" and "k" is a "way of classification" or vice versa, or (ii) "i", "j", and "k" are all "ways of classification". The obvious modifications in the cases (i) and (ii) will be in the fact that the appropriate p's will have carets over them and the rest will not.

8. "No interaction" in a 2 x 2 x 2 table. Consider in this case the hypothesis (4.2.5), and write it out in full as follows:

$$(8.1) \qquad H_0: \; p_{111} = \frac{q_{110}q_{101}q_{011}}{q_{100}q_{010}q_{001}}, \quad p_{211} = \frac{q_{210}q_{201}q_{011}}{q_{200}q_{010}q_{001}},$$

$$p_{112} = \frac{q_{110}q_{102}q_{012}}{q_{100}q_{010}q_{002}}, \quad p_{212} = \frac{q_{210}q_{202}q_{012}}{q_{200}q_{010}q_{002}},$$

$$p_{121} = \frac{q_{120}q_{101}q_{021}}{q_{100}q_{020}q_{001}}, \quad p_{221} = \frac{q_{220}q_{201}q_{021}}{q_{200}q_{020}q_{001}},$$

$$p_{122} = \frac{q_{120}q_{102}q_{022}}{q_{100}q_{020}q_{002}}, \quad p_{222} = \frac{q_{220}q_{202}q_{022}}{q_{200}q_{020}q_{002}}.$$

It is easy to check that by eliminating the q's, we have, what we will call, the "no interaction" constraints, which, in this case, represent just one relation among the p's, namely

$$(8.2) \qquad \frac{p_{111}p_{221}}{p_{211}p_{121}} = \frac{p_{112}p_{222}}{p_{212}p_{122}}.$$

There is, of course, the other side condition on the p's:

$$(8.3) \qquad \sum_{i,j,k} p_{ijk} = 1.$$

Recalling again section 2, the likelihood function can be written as

$$(8.4) \qquad \phi \sim \prod_{i,j,k} p_{ijk}^{\,n_{ijk}}.$$

Now consider

Case I. "i", "j", and "k" all are "variates". The problem is to estimate the

p_{ijk}'s by maximizing ϕ subject to the constraints (8.2) and (8.3). Introducing the usual Lagrangian multipliers on (8.2) and (8.3) we have the maximum likelihood equations

$$\text{(8.5)} \qquad \frac{n_{ijk}}{p_{ijk}} + \frac{\lambda}{p_{ijk}} + \mu = 0 \qquad (ijk = 111,\ 221,\ 212,\ 122)$$

$$\frac{n_{ijk}}{p_{ijk}} - \frac{\lambda}{p_{ijk}} + \mu = 0 \qquad (ijk = 112,\ 222,\ 211,\ 121) \quad .$$

Now multiplying by p_{ijk}, and summing over i,j,k, and using (8.3), we have $\mu = -n$, and

$$\text{(8.6)} \qquad \hat{p}_{ijk} = +(n_{ijk} + \lambda) / n \qquad (ijk = 111,\ 221,\ 212,\ 122)$$

$$\hat{p}_{ijk} = +(n_{ijk} - \lambda) / n \qquad (ijk = 112,\ 222,\ 211,\ 121).$$

Substituting in (8.2), we have for λ the cubic equation

$$\text{(8.7)} \qquad \frac{(n_{111} + \lambda)(n_{221} + \lambda)}{(n_{211} - \lambda)(n_{121} - \lambda)} = \frac{(n_{112} - \lambda)(n_{222} - \lambda)}{(n_{212} + \lambda)(n_{122} + \lambda)} \quad .$$

Solving for λ and substituting in (8.6) we have the estimated $\hat{p}_{ijk}$'s occurring in the usual χ^2. Since

$$\text{(8.8)} \qquad n_{ijk} - n\hat{p}_{ijk} = -\lambda \qquad (ijk = 111,\ 221,\ 212,\ 122)$$

$$n_{ijk} - n\hat{p}_{ijk} = +\lambda \qquad (ijk = 112,\ 222,\ 211,\ 121) \ ,$$

the final χ^2 is given by

$$\text{(8.9)} \qquad \chi^2 = \frac{\lambda^2}{n} \sum_{i,j,k=1}^{2} \hat{p}_{ijk}^{\,-1}.$$

This will be a χ^2 with d.f. = the total number of cells (8 here) - [the apparent number of parameters (8 here) - the number of "no interaction" constraints (1 , here) - the number of linear relations on the p's coming from the linear constraints on the n's (1 here)] - [the number of linear relations on the n's (1 here)] = the number of "no interaction" constraints = 1, in this case. It is easy to see from this that, in all cases, no matter whether "i", "j", and "k" are all "variates", or some are "variates" and some "ways of classification", or all are "ways of classification", we are going to end up with a χ^2 with d.f. exactly equal to the number of "no interaction" constraints like (8.2).

Notice that in (8.5), the Lagrangian μ goes with the constraint $\sum_{i,j,k} p_{ijk} = 1$ which stems from $\sum_{i,j,k} n_{ijk} = n$, and the Lagrangian λ goes with the "no interaction" constraints (8.2).

<u>Case II.</u> <u>"i" and "j" are "ways of classification, and "k" is a "variate"</u>. This is the case of a two-way analysis of variance, and it is clear from the remarks after (8.9) that we shall end up with a χ^2 with the same d.f. = 1. Here

(8.10) $$p_{ijo} = \sum_k p_{ijk} = n_{ijo}/n \quad \text{(fixed)} \quad \text{with } i, j = 1, 2.$$

The maximum likelihood equations for the p's subject to (8.2) and (8.10) will now be

(8.11)
$$\frac{n_{ijk}}{p_{ijk}} + \frac{\lambda}{p_{ijk}} + \mu_{ij} = 0 \qquad (ijk = 111, 221, 212, 122)$$
$$\frac{n_{ijk}}{p_{ijk}} - \frac{\lambda}{p_{ijk}} + \mu_{ij} = 0 \qquad (ijk = 112, 222, 211, 121).$$

Consider any (ij). Take say (11), and notice that λ goes with $k = 1$, and $-\lambda$ with $k = 2$, so that multiplying by p_{ijk} and summing over $k = 1$ and 2, we have

(8.12) $$n_{ijo} + \mu_{ij}p_{ijo} = 0 \quad \text{or} \quad \mu_{ij} = -n, \text{ using (8.10)}.$$

Hence substituting from (8.11) in (8.2), we have the same equation in λ, and finally the same $\hat{p}$'s, and thus the same χ^2 as in the previous Case I.

Case III. "i" is a "way of classification", and "j" and "k" are "variates". Again from the remarks after (8.9) we observe that we will get a χ^2 with the same d.f. = 1. Here

$$p_{ioo} = n_{ioo}/n \quad \text{(fixed), with } i = 1, 2. \tag{8.13}$$

The maximum likelihood equations for the p's, subject to (8.2) and (8.13) will now be

$$\begin{aligned} \frac{n_{ijk}}{p_{ijk}} + \frac{\lambda}{p_{ijk}} + \mu_i &= 0 \qquad (ijk = 111,\ 221,\ 212,\ 122) \\ \frac{n_{ijk}}{p_{ijk}} - \frac{\lambda}{p_{ijk}} + \mu_i &= 0 \qquad (ijk = 112,\ 222,\ 211,\ 121). \end{aligned} \tag{8.14}$$

Notice that for a given i, say 1, we have λ with jk = 11, 22, and $-\lambda$ with jk = 12, 21, so that if we multiply by p_{ijk} and sum over jk, we will have

$$n_{ioo} + \mu_i p_{ioo} = 0 \qquad \text{or} \qquad \mu_i = -n, \quad \text{using (8.13)}. \tag{8.15}$$

Hence substituting from (8.14) in (8.2) we have the same equations in λ and finally the same $\hat{p}$'s, and thus the same χ^2 as in the previous Cases I and II.

Case IV. "i", "j", and "k" all are "ways of classification" [1]. We shall get a χ^2 with the same d.f. = 1. Here

$$p_{ijo} = n_{ijo}/n \text{ (fixed)}; \quad p_{iok} = n_{iok}/n \text{ (fixed)}; \quad p_{ojk} = n_{ojk}/n \quad \text{(fixed)} \tag{8.16}$$

with i,j,k = 1,2. Notice that the relations in (8.16) are not independent. In fact, from one angle it will be seen that if we put $p_{111} = x$ (say), then all the other p's will be completely given in terms of x and the fixed marginals of (8.16). Then substituting in (8.2) we can find x. From another angle (which should, of course, finally give the same result) we notice, by putting $n_{111} - np_{111} = x$ and using (8.16), that

$$(8.17)\qquad \begin{aligned} n_{ijk} - np_{ijk} &= -x \qquad \text{for} \quad (ijk = 221,\ 212,\ 122) \\ n_{ijk} - np_{ijk} &= +x \qquad \text{for} \quad (ijk = 112,\ 222,\ 211,\ 121). \end{aligned}$$

This means that x is exactly λ of (8.5) or (8.11) or (8.14), so that we have the same equations in x as we had in λ in the previous cases. Hence we have the same expressions for p_{ijk} in terms of n_{ijk} as we had for $\hat{p}_{ijk}$ in terms of n_{ijk} in the previous case. And hence we have finally the same χ^2 in terms of the n_{ijk}'s as in the previous cases.

9. "No interactions" in an r x s x t table. Let us consider here the hypothesis of "no interaction", and try to eliminate the q's. To fix our ideas, consider first the case of a 2 x 2 x t table. Looking into the mechanics by which (8.2) is obtained from (8.1), it is easy to see that, corresponding to (8.2) we are going to have

$$(9.1)\qquad \frac{p_{11t}p_{22t}}{p_{21t}p_{12t}} = \frac{p_{11,t-1}p_{22,t-1}}{p_{21,t-1}p_{12,t-1}} = \frac{p_{11,t-2}p_{22,t-2}}{p_{21,t-2}p_{12,t-2}} = \cdots = \frac{p_{111}p_{221}}{p_{211}p_{121}} .$$

For a general r x s x t table we can figure out that we are going to have the following "no interaction" constraints:

$$(9.2)\qquad \frac{p_{rst}p_{ijt}}{p_{ist}p_{rjt}} = \frac{p_{rsk}p_{ijk}}{p_{isk}p_{rjk}} , \qquad \text{for} \qquad \begin{aligned} k &= 1,\ 2,\ \ldots,\ (t-1) \\ j &= 1,\ 2,\ \ldots,\ (s-1) \\ i &= 1,\ 2,\ \ldots,\ (r-1). \end{aligned}$$

This gives us (t-1)(s-1)(r-1) constraints on the p_{ijk}'s. Checking the mechanics of the derivation of (8.2) from (8.1), it will be seen that (9.2) yields a set of independent and exhaustive relations among the p's by eliminating the q's from (4.2.5). Here p_{rst} is, as it were, a pivotal element, and r, s, and t the pivotal subscripts. We can make any other three subscripts the pivotal ones, and thus obtain another set of independent and exhaustive relations like (9.2) which would be exactly equivalent to (9.2), and so on.

Our likelihood function is

$$\phi \sim \prod_{i,j,k} p_{ijk}^{n_{ijk}} . \tag{9.3}$$

Looking into the mechanics by which, in the case of a 2 x 2 x 2 table, the four cases, namely (i) "i", "j", and "k" all being "variates", (ii) any one being a "variate" and the other two "ways of classification", (iii) any two being "variates" and the remaining one a "way of classification", and (iv) all being "ways of classification", were shown to be mathematically equivalent (in terms of testing by χ^2), we can verify that this will hold for the general r x s x t table also. In all the cases, we have the same form of χ^2, distributed with d.f. (r-1)(s-1)(t-1). We discuss, therefore, only

Case I. All are "variates". Here we have to maximize (9.3) subject to the "no interaction" constraints (9.2), and the further constraint

$$\sum_{i,j,k} p_{ijk} = 1. \tag{9.4}$$

Introducing for (9.1) the Lagrangian multipliers λ_{ijk} [i = 1, 2, ..., (r-1); j = 1, 2, ..., (s-1); k = 1, 2, ..., (t-1)], and for (9.4) the Lagrangian multiplier μ, and maximizing (9.3), we have for $\hat{p}_{ijk}$ the typical equations

$$(9.5) \qquad \frac{n_{rst}}{p_{rst}} + \frac{\sum_{i=1}^{(r-1)} \sum_{j=1}^{(s-1)} \sum_{k=1}^{(t-1)} \lambda_{ijk}}{p_{rst}} + \mu = 0$$

$$\frac{n_{ist}}{p_{ist}} - \frac{\sum_{j=1}^{(s-1)} \sum_{k=1}^{(t-1)} \lambda_{ijk}}{p_{ist}} + \mu = 0$$

$$\frac{n_{rjt}}{p_{rjt}} - \frac{\sum_{i=1}^{(r-1)} \sum_{k=1}^{(t-1)} \lambda_{ijk}}{p_{rjt}} + \mu = 0$$

$$\frac{n_{rsk}}{p_{rsk}} - \frac{\sum_{i=1}^{(r-1)} \sum_{j=1}^{(s-1)} \lambda_{ijk}}{p_{rsk}} + \mu = 0$$

$$\frac{n_{ijt}}{p_{ijt}} + \frac{\sum_{k=1}^{(t-1)} \lambda_{ijk}}{p_{ijt}} + \mu = 0$$

$$\frac{n_{isk}}{p_{isk}} + \frac{\sum_{j=1}^{(s-1)} \lambda_{ijk}}{p_{isk}} + \mu = 0$$

$$\frac{n_{rjk}}{p_{rjk}} + \frac{\sum_{i=1}^{(r-1)} \lambda_{ijk}}{p_{rjk}} + \mu = 0$$

$$\frac{n_{ijk}}{p_{ijk}} - \frac{\lambda_{ijk}}{p_{ijk}} + \mu = 0$$

with, of course, $i = 1, 2, \ldots, (r-1)$; $j = 1, 2, \ldots, (s-1)$; $k = 1, 2, \ldots, (t-1)$. Notice that with the pivotal subscripts (rst) goes a triple summation over the λ's and a positive sign before that expression; with just one subscript changed goes a double summation over the λ's and a negative sign before that expression; with two of the subscripts changed goes a single summation over the λ's and a positive sign before that expression; and finally with all the subscripts changed, we have a single λ_{ijk} with a negative sign before it.

As in the case of the 2 x 2 x 2, it is easy to see by multiplying both sides of (9.5) by p_{ijk} and summing over i,j,k, that $\mu = -n$. Thus solving for the p_{ijk}'s in terms of the n_{ijk}'s and λ_{ijk}'s, and substituting in the "no interaction" constraints (9.2), we have for λ_{ijk} the following equations [for i = 1, 2, ..., (r-1); j = 1, 2, ..., (s-1); k = 1, 2, ..., (t-1)]:

$$(9.10) \qquad \frac{(n_{rst} + \mu_{rst})\,(n_{ijt} + \mu_{ijt})}{(n_{ist} - \mu_{ist})\,(n_{rjt} - \mu_{rjt})} = \frac{(n_{rsk} - \mu_{rsk})\,(n_{ijk} - \mu_{ijk})}{(n_{isk} + \mu_{isk})\,(n_{rjk} + \mu_{rjk})}$$

where μ_{rst} stands for the triple summation expression in (9.5), μ_{ist}, μ_{rjt}, μ_{rsk} for the double summation expressions in (9.5), μ_{ijt}, μ_{isk}, μ_{rjk} for the single summation expressions in (9.5), and μ_{ijk} is simply λ_{ijk}. Solving equations (9.10) for the μ_{ijk}'s, and ultimately for the λ_{ijk}'s, in terms of the n_{ijk}'s, we can find the $\hat{p}_{ijk}$'s. Substituting these values in the usual expression for χ^2 we have

$$(9.11) \qquad \sum_{i,j,k} \mu_{ijk}^2 \,/\, (n_{ijk} + \eta_{ijk}\mu_{ijk})$$

where i = 1, 2, ..., r; j = 1, 2, ..., s; k = 1, 2, ..., t; and where

$\eta_{ijk} = +1$ if ijk = rst (the pivotal subscript);

$\eta_{ijk} = -1$ if any one subscript differs from the corresponding pivotal subscript;

$\eta_{ijk} = +1$ if any two subscripts differ from the corresponding pivotal subscripts;

$\eta_{ijk} = -1$ if all subscripts differ from the corresponding pivotal subscripts.

The method of solving equations (9.10) for the μ_{ijk}'s and finally for the λ_{ijk}'s in terms of the n_{ijk}'s on modern high speed computers will be discussed in a later paper.

10. Concluding remarks. The extension from tables of three to more than three dimensions does not bring up any new problems so far as concepts like "no multiple correlation", "no partial correlation", etc. are concerned. A new feature with, for example, a four-way table would be the concept of "no correlation"between "(ij)" and "(kℓ)" which can be expressed as H_0: $p_{ijk\ell} = p_{ijoo}p_{ook\ell}$, and tested in a straightforward manner. The generalization to any number of dimensions of the concept of independence between two sets of "variates" is obvious. Replacement of some of the "variates" by "ways of classification" will only make the final interpretation a little different. The concept of "no interaction", however, is not one of trivial generalization. In a four-way table the hypothesis analogous to (4.2.5), that is, the hypothesis of "no second order interaction", seems to be

$$(10.1) \qquad H_0: \quad p_{ijk\ell} = \frac{q_{ijko}q_{ijo\ell}q_{iok\ell}q_{ojk\ell}q_{iooo}q_{ojoo}q_{ooko}q_{ooo\ell}}{q_{ijoo}q_{ioko}q_{ioo\ell}q_{ojko}q_{ojo\ell}q_{ook\ell}} \quad .$$

In this case the hypotheses of four separate "no first-order interactions" follow exactly the same pattern as in section 9, and need not be separately considered. The extension of (10.1) to higher order "no interactions", in the case of tables of higher dimensions, forms a certain pattern which has been worked out and which will be discussed in a later paper. The technique of testing (10.1) and "no interaction "hypotheses of higher order is essentially similar, in principle, to what has been discussed in section 9. The details alone are more complicated. For higher order "no interactions" there are, however, various intermediate cases of considerable interest which will be discussed later.

We have discussed the hypotheses of "no multiple correlation", "no partial correlation", etc., but have not introduced any measures of "multiple correlation", "partial correlation", "interaction", etc. This is now under investigation. One measure might be the expected value of χ^2 when the hypothesis is not true minus the expected value when the hypothesis is true (which is a simple and well-known expression). No exact tests, with some reasonably good properties against the class of relevant alternatives (to the hypotheses), have been discussed here, nor have the powers against permissible alternatives, of the X^2-test been discussed. In several of the cases, the power functions would be easily available from previous work by others, but in other cases, they would have to be worked out. These will be discussed later.

We give below only six references. The sources we have drawn upon most are [1], [4], [5], [6]. For a critical review of much of the previous work on the subject, and a reasonably exhaustive bibliography we would recommend [2] and [3]. The reader will perceive that this paper has some (but not much) overlap with [2] in the general sector of "independence" in a two-way table.

REFERENCES

[1] Bartlett, M. S., "Contingency table interactions," Jour. Roy. Stat. Soc. Suppl., Vol. 2 (1935), pp. 248-252.

[2] Cochran, W. G., "The X^2-test of goodness of fit," Annals of Math. Stat., Vol. 23 (1952), pp. 315-345.

[3] Cochran, W. G., "Some methods for strengthening the common X^2-tests," Biometrics, Vol. 10 (1954), pp. 417-451.

[4] Cramer, H., Mathematical Methods of Statistics, Princeton University Press, 1946, Chap. 30.

[5] Fisher, R. A., "On the interpretation of chi-square from contingency tables, and the calculation of P," Jour. Roy. Stat. Soc., Vol. 85 (1922), pp. 87-94.

[6] Yule, G. Udny and Kendall, M. G., An Introduction to the Theory of Statistics, Hafner Publishing Company, 1950, Chapter 1 - 4.