Professor Miss G. Cox.

SOME NONPARAMETRIC ANALOGUES OF "NORMAL" ANOVA, MANOVA, AND OF STUDIES IN "NORMAL" ASSOCIATION

by

S. N. Roy and V. P. Bhapkar

University of North Carolina

This research was supported partly by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 215
January, 1959

SOME NONPARAMETRIC ANALOGUES OF "NORMAL" ANOVA, MANOVA, AND OF STUDIES IN "NORMAL" ASSOCIATION[1]

By S. N. Roy and V. P. Bhapkar

University of North Carolina

Summary. We imagine the experimental data to be given in the form of frequencies in cells determined by a (finitely) multiway cross classification, with predefined categories, finite in number, along each way of classification. According as the marginal frequencies along any way or dimension are held fixed or left free, that dimension or "way" will be said to be associated with a "factor" or a "response." The general probability model is that of a product of several multinomial distributions, the "probabilities over the different cells" being supposed to be unknown, to start with, except for certain broad equality and inequality relations. According as a set of (real) values is or is not associated with the categories along any "way of classification" (with or without fixed marginal frequencies) that "way of classification" will be said to be a structured or unstructured response or factor. It is the purpose of this paper (which is a follow-up of previous papers [1,2,3,4,5]) to pose hypotheses that the authors consider to be generalizations appropriate to this set-up of the usual hypotheses in

---

[1]This research was supported partly by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

classical "normal" univariate "fixed effects" analysis of variance or ANOVA, "normal" multivariate "fixed effects" analysis of variance or MANOVA, and in analysis of various kinds of "normal" independence and to offer large sample tests for such hypotheses. Point estimation and confidence bounds on meaningful parametric functions would follow on the lines discussed in the previous papers [1,2,4,5] and are not separately discussed here. Nor are more general decision procedures (e.g., ranking, choosing the best, etc.) that are no doubt physically more meaningful than the usual testing of hypotheses or even estimation, discussed in this paper. Another severe restriction is this. For each (kind of) response or factor, we make a distinction between the structured and the unstructured case; but even for structured factors we give the pre-Box treatment, that avoids the realistic problem of exploring the nature of the dependence of the response on the structured factor or factors. Such progress as has been recently made beyond these restrictions will be reported in subsequent papers. Part of the material on responses that are categorical without a structure presented here is a recapitulation of earlier discussion [1,2,4,5] while the rest is new, the material on responses that are categorical with a structure being all new.

1. Introduction and preliminaries. Let

$$\Phi = \prod_j \left[\frac{n_{oj}!}{\prod_i n_{ij}!} \prod_i p_{ij}^{n_{ij}} \right] \tag{1.1}$$

denote a product-multinomial distribution of a certain number of experimental units such that $\sum_j p_{ij} \equiv p_{oj} = 1$ and $\sum_i n_{ij} = n_{oj}$ is fixed with $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,s$. In other words, the set-up might be regarded as one of stratified sampling in which "j" refers to a stratum and "ij" to the i-th category in the j-th stratum, n_{oj} the preassigned sample size from the j-th stratum, n_{ij} the observed frequency of experimental units in the i-th category of the j-th stratum, and p_{ij} the probability of having an experimental unit from the j-th stratum in the i-th category. It should be further noticed that i may be a multiple subscript, say, $i_1 i_2 \ldots i_k$ with $i_1 = 1,2,\ldots,r_1$; $i_2 = 1,2,\ldots,r_2$; $\ldots$; $i_k = 1,2,\ldots,r_k$, so that, all combinations being supposed to be allowed, $r = r_1 r_2 \ldots r_k$. Likewise, j also might be a multiple subscript, say, $j_1 j_2 \ldots j_\ell$ with $j_1 = 1,2,\ldots,s_1$; $j_2 = 1,2,\ldots,s_2$; $\ldots$; $j_\ell = 1,2,\ldots,s_\ell$, but with this important distinction that all combinations may not be allowed. In other words, the classification with respect to the j's might be based on a multi-dimensional incomplete design such that, given $j_1,\ldots,j_{\ell-1}$, $j_\ell \in (s_\ell)_{j_1 j_2 \ldots j_{\ell-1}}$ (j_ℓ takes a set of values which is a subset of the integers $1,2,\ldots,s_\ell$, depending upon $j_1 j_2 \ldots j_{\ell-1}$), and similarly, $j_{\ell-1} \in (s_{\ell-1})_{j_1 j_2 \ldots j_{\ell-2}}, \ldots;$ $j_2 = (s_2)_{j_1}$ and $j_1 = 1,2,\ldots,s_1$. Thus s is not necessarily $s_1 s_2 \ldots s_\ell$, but can be computed from this characterization of the incomplete design. This will be called a k-response (or k-variate) and ℓ-factor problem, "i_1","i_2",...,"i_k" being called responses (or variates) and "j_1","j_2",...,"j_ℓ" being called factors. Summation over a subscript will be indicated by replacing that subscript by o; on n such a summation will, of course, indicate a marginal frequency. On p, however, such a summation will indicate a marginal probability only when the subscript in question is associated with a response, not when it is associated with a factor. It may be noticed that n_{oj} in (1.1) and p_{oj} a few lines below really stand for $n_{o\ldots oj_1\ldots j_\ell}$ and $p_{o\ldots oj_1\ldots j_\ell}$. To fix our ideas, consider, for example, three subscripts, say $i_1 i_2 i_3$. Then a typical quantity (other than n or p) depending upon $i_1 i_2 i_3$ will be denoted by, say, $q_{i_1 i_2 i_3}$. If it depends on i_1 and i_3 but not i_2 it will be denoted by $q_{i_1 * i_3}$, and so on. Thus a star in place of a subscript will indicate that the quantity in question is independent of that subscript.

Next, as to the hypothesis, it may be noticed that (1.1) assumes constraints on p_{ij}'s, namely, that $\sum_i p_{ij} = p_{oj} = 1$, aside from inequalities of the type discussed in [1,2,4,5]. If, with (1.1) as the model, a hypothesis $\mathcal{H}_1$ is given in the form of certain constraints on p_{ij}'s (in addition, of course, to $\sum_i p_{ij} = 1$), then the large sample test of $\mathcal{H}_1$ under (1.1) for model is in terms of a statistic given by

$$(1.2) \qquad \sum_{i,j} \frac{(n_{ij} - n_{oj}\hat{p}_{ij})^2}{n_{oj}\hat{p}_{ij}} ,$$

in which $\hat{p}_{ij}$'s are estimates (in terms of n_{ij}'s) of p_{ij}'s which maximize the Φ of (1.1) subject to $\sum_i p_{ij} = 1$ and to the further constraints on p_{ij}'s that define the hypothesis. This statistic, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, is distributed as a χ^2 with degrees of freedom = the no. of <u>independent</u> constraints on p_{ij}'s (beyond $\sum_i p_{ij} = 1$) that define the hypothesis. It may be observed that in place of the maximum likelihood estimates of p_{ij}'s, we might as well have considered the minimum χ^2 estimates or any set of estimates belonging to the broader well-known class of BAN estimates. Likewise, instead of the statistic (1.2) we might as well have considered the slightly different one known as χ_1^2 (with the same asymptotic χ^2-distribution). At this stage we can go further, introduce some additional constraints on p_{ij}'s and thus define a new hypothesis $\mathcal{H}_2 \subset \mathcal{H}_1 \subset (1.1)$. The test of $\mathcal{H}_2$, under (1.1) for model, is in terms of a statistic given by

$$(1.3) \qquad \sum_{i,j} \frac{(n_{ij} - n_{oj}\hat{\hat{p}}_{ij})^2}{n_{oj}\hat{\hat{p}}_{ij}} ,$$

which, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, has the χ^2-distribution with degrees of freedom = no. of independent constraints on p_{ij}'s (beyond $\sum_i p_{ij} = 1$)

that define $\mathcal{H}_1$ plus the additional ones that define $\mathcal{H}_2$. However, if we want to test $\mathcal{H}_2$ under $\mathcal{H}_1$ for model, then the test will be given in terms of a statistic [1,2]

$$(1.4) \qquad \sum_{i,j} \frac{(n_{ij} - n_{oj}\hat{\hat{p}}_{ij})^2}{n_{oj}\hat{\hat{p}}_{ij}} - \sum_{i,j} \frac{(n_{ij} - n_{oj}\hat{p}_{ij})^2}{n_{oj}\hat{p}_{ij}},$$

which, in the limit as $n \to \infty$ subject to n_{oj}/n's being held fixed, has the [1,2,4,5] χ^2-distribution with degrees of freedom = no. of free parameters under the model - no. of free parameters under the hypothesis = the no. of additional independent constraints on the p_{ij}'s that define $\mathcal{H}_2$ under $\mathcal{H}_1$ for model. The asymptotic power of such tests for $\mathcal{H}_1$ under (1.1), $\mathcal{H}_2$ under (1.1), and $\mathcal{H}_2$ under $\mathcal{H}_1$ have been defined and obtained (in an abstract form) in [1,2]. So far as these large sample tests (as also their asymptotic power) are concerned, the treatment given in an abstract form in [1,2,4,5] is perfectly general and can take care of both situations--namely when the responses (or variates) and/or the factors are structured or unstructured.

Now let us consider the case of a structured response. A response, say "i_m", will be said to be structured if there is a set of (real) values $a^{(m)}_{i_m}$ $(i_m = 1,2,\ldots,r_m)$ associated as weights with the set of categories "i_m". (1) If the categories "i_m" are class intervals (numbered in the proper order) for a continuous stochastic variate, then the distances from an arbitrary origin of, say, the mid-points of such intervals form a natural set of $a_{i_m}^{(m)}$'s. (2) Likewise, if, back of us, there is a discrete (real) valued stochastic variate taking on values $a_1^{(m)} \leq a_2^{(m)} \leq \dots \leq a_{r_m}^{(m)}$, then these might be the natural weights to go with the categories in that direction, which also might be defined and numbered in a natural manner. (3) If the response is not even a discrete (real) valued variate but is categorical with an implied ranking (like good, fair, poor, etc.) then the categories can be defined and numbered in a natural way and scores can be assigned to these categories. (4) Finally, if the response be in terms of categories without even an implied ranking (for example, like black, yellow, red, etc.), then the numbering of these categories is purely nominal and there is no natural system of weights to go prima facie with the categories. However, on economic or other considerations we might assign a system of (real-valued) weights to these categories in which the initially unstructured response would become structured. But even in the case of (1), (2) and (3), such considerations might induce a system of weights different from the ones that are "natural" for (1), (2) or (3). In most realistic problems, especially of a utilitarian (rather than purely scientific) nature, each kind of response eventually becomes structured although it may be unstructured to start with. On the other hand, so far as the factors are concerned, any factor could well be either continuous or discrete or categorical with an implied ranking or purely categorical. Thus, in general, in a complex realistic situation some of the responses are of type (1), some of type (2), some of type (3) and the rest of type (4), with a similar division into four types for the factors. In the present paper we shall discuss three types of problems, namely (i) where all responses are unstructured and so also all factors, (ii) some responses are structured and the rest are unstructured and all factors are unstructured, and (iii) some responses are structured, the rest unstructured, and some factors are structured while the rest are unstructured. So far as the problems of interest are concerned the main distinction between the unstructured and the structured case seems to be this. It is feasible and also useful, in the structured case, to go ahead and define certain overall (or, to borrow a phrase from differential geometry, im grössen) aspects of the distribution which it is not possible to define for the unstructured case. Then, while in the structured case it is still possible to study the same problems as in the unstructured case, such problems (interesting for the unstructured case) are hardly of the same interest for the structured case, and the problems involving the newly-defined overall aspects are the ones that become more meaningful. Problems involving other possibilities will be discussed in later papers.

To make everything concrete we shall consider in this paper only some three-way, four-way and five-way tables.

For the three-way table (ijk) we shall consider the following cases. (a) "i," "j", "k" all responses and all unstructured or one or more unstructured and the rest structured, (b) "i" an unstructured response and "j" and "k" unstructured factors, (c) "i" is a structured response and "j" and "k" are factors, either both unstructured or one structured and the other unstructured. For four-way ($ijk\ell$) tables we consider problems similar to (a) above; and for five-way ($ijk\ell m$) tables we consider the cases where (a) "i" and "j" are unstructured responses and "k", "ℓ" and "m" are unstructured factors, (b) "i" and "j" are structured responses while "k", "ℓ" and "m" are either all unstructured factors or "k" is a structured, and "ℓ" and "m" are unstructured factors. This will serve for illustration, and what happens as we increase the dimensionality of the table (but otherwise stay within the limitations indicated earlier) will be immediately obvious.

2. <u>A three-way table (ijk) in which "i", "j" and "k" all are responses or variates</u>. Let us consider a single-multinomial distribution

$$\Phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}} , \tag{2.1}$$

where $\sum_{i,j,k} p_{ijk} = 1$ and $\sum_{i,j,k} n_{ijk} = n$ (fixed).

2.1. <u>The case where "i", "j" and "k" are all unstructured</u>.--In this set-up we shall consider, for illustration, some important specific hypotheses, which are as follows.

$$(2.1.1) \qquad \mathcal{H}_{01}: \; p_{ijk} = \frac{p_{iok}p_{ojk}}{p_{ook}} \quad ,$$

which can be described as the hypothesis that "i" and "j" are independent in the conditional distribution of "i" and "j", given "k", or, in other words, that there is no partial association between "i" and "j", given "k." There are two equivalent ways to get at (2.1.1). One is to notice that the conditional joint distribution of "i" and "j", given "k", is p_{ijk}/p_{ook} and the conditional marginal distribution of "i", given "k", is p_{iok}/p_{ook}, that of "j", given "k", is p_{ojk}/p_{ook}, which leads to (2.1.1) as the condition that there is no partial association between "i" and "j". Another way is to start with the condition

$$(2.1.2) \qquad p_{ijk}/p_{ojk} \text{ is independent of "j"} = q_{i*k} \quad \text{(say)},$$

which means that the conditional distribution of "i", given "j" and "k", is independent of "j", rewrite (2.1.2) in the form

$$(2.1.3) \qquad p_{ijk} = p_{ojk}q_{i*k} \quad ,$$

sum up both sides over "j" and thus verify that (2.1.2) $\Longleftrightarrow$ (2.1.1). This hypothesis and the χ^2 appropriate to it has already been discussed in [2,4], except that the approach (2.1.2) was not considered.

We proceed next to the hypothesis that "j" and "k" are independent and so also "i" and "k", or in symbols, (2.1.4) $\qquad \mathcal{H}_{02}$: $p_{ojk} = p_{ojo}\,p_{ook}$ and $p_{iok} = p_{ioo}p_{ook}$,

and note that $\mathcal{H}_{01} \cap \mathcal{H}_{02} \Longrightarrow$

(2.1.5) $\qquad \mathcal{H}_{03}$: $p_{ijk} = p_{ioo}p_{ojo}p_{ook}$,

which is the hypothesis of overall independence of "i", "j" and "k". It should also be noticed that $\mathcal{H}_{02} \not\Longrightarrow$

(2.1.6) $\qquad \mathcal{H}_{04}$: $p_{ijk} = p_{ijo}p_{ook}$,

which is the hypothesis that "k" is multiply independent of "ij", or, in other words, that there is no multiple association between "ij" and "k". It is easy to check by summing the two sides of (2.1.6) over "i" and over "j" separately that $\mathcal{H}_{04} \Longrightarrow \mathcal{H}_{02}$, but not, in general, the other way around (unlike what happens in the multivariate normal case). It is also easy to check that one could also get to (2.1.6) by either starting from

(2.1.7) $\qquad p_{ijk}/\,p_{ijo}$ is independent of "ij" $= q_{**k}$ (say)

or from

(2.1.8) $\qquad p_{ijk}/\,p_{ook}$ is independent of "k" $= q_{ij*}$ (say).

Another hypothesis which might be of interest is that of two by two independence of "i", "j" and "k", or in symbols,

(2.1.9) $\qquad \mathcal{H}_{05}$: $p_{ijo} = p_{ioo}p_{ojo}$, $\;p_{iok} = p_{ioo}p_{ook}$ and

$$p_{ojk} = p_{ojo}p_{ook} .$$

It is easy to check that $\mathcal{H}_{03} \Longrightarrow \mathcal{H}_{05}$ but unlike what happens in the multivariate normal case it is not here (i.e., in general) the other way around, which is, of course, well known to probabilists. We observe that the experimenter might be very well interested in $\mathcal{H}_{03}$ but not necessarily in $\mathcal{H}_{04}$, and likewise in $\mathcal{H}_{05}$ but not necessarily in $\mathcal{H}_{03}$.

2.2 <u>The case where "i" is structured but not necessarily "j" or "k"</u>.--Let us start by considering a hypothesis which does not have any analogue under the previous set-up, namely,

(2.2.1) $\mathcal{H}_{00}$: $\sum_i a_i p_{ijk} / p_{ojk}$ = an unknown function of "j" $\times$ an unknown function of "k" $= q_{*j*} q_{**k}$ (say)

or

(2.2.2) $\mathcal{H}^*_{00}$: $\sum_i a_i p_{ijk} / p_{ojk} = q^*_{*j*} + q^*_{**k}$ (say) .

We next consider a hypothesis analogous to $\mathcal{H}_{01}$. We consider the version of $\mathcal{H}_{01}$ given by (2.1.2), and express this corresponding hypothesis in this set-up by

(2.2.3) $\mathcal{H}^*_{01}$: $\sum_i a_i p_{ijk} / p_{ojk}$ is independent of "j" $= q_{**k}$ (say),

where a_i's are supposed to be given. This is, of course, a special case of (2.2.1). By rewriting (2.2.3) as $\sum_i a_i p_{ijk} = p_{ojk} q_{**k}$ and summing both sides over "j" it is easy to check that (2.2.3) could be rewritten in the equivalent form

(2.2.4) $\mathcal{H}^*_{01}$: $\sum_i a_i [p_{ijk} - \frac{p_{iok} p_{ojk}}{p_{ook}}] = 0$.

It should be noticed that $\mathcal{H}_{01} \Longrightarrow \mathcal{H}^*_{01}$ but not the other way around, for a given set of a_i's. However, if (2.2.4) is to hold for sets of a_i's, then (2.2.4) (for all sets of a_i's) $\Longleftrightarrow \mathcal{H}_{01}$.

Also analogous to $\mathcal{H}_{02}$ (but with an interchange of subscripts) we now have

(2.2.5) $\mathcal{H}^*_{02}$: $\sum_i a_i p_{ijo} / p_{ojo}$ is independent of "j" $= \lambda$ (say)

and $\sum_i a_i p_{iok} / p_{ook}$ is independent of "k" $= \mu$ (say),

leading eventually to

(2.2.6) $\mathcal{H}^*_{02}$: $\sum_i a_i (p_{ijo} - p_{ioo} p_{ojo}) = 0$ and

$$\sum_i a_i (p_{iok} - p_{ioo} p_{ook}) = 0 ,$$

the same remarks being applicable as after (2.2.4). Likewise, analogous to $\mathcal{H}_{04}$ (but with an interchange of subscripts) we have

(2.2.7) $\mathcal{H}^*_{04}$: $\sum_i a_i p_{ijk} / p_{ojk}$ is independent of "jk" $= \nu$ (say),

leading to

(2.2.8) $\mathcal{H}^*_{04}$: $\sum_i a_i (p_{ijk} - p_{ioo} p_{ojk}) = 0$,

the same remarks being applicable again as after (2.2.4). The analogues of $\mathcal{H}_{03}$ and $\mathcal{H}_{05}$ are obvious and can be left to the reader. The more interesting case is where more than one of "i", "j" and "k" are structured.

2.3 <u>The case where "i" and "j" are structured but not necessarily "k".</u>--If, for concreteness, we fix on "i" (say), then the hypothesis of independence of "i" with respect to "j" and "k" of the types (2.2.3), (2.2.5) and (2.2.7) (as also the tests) will stay the same as before. If these are tested and rejected, then (in subsequent experiments with the same type of material) we have other interesting and useful possibilities. Assuming a set of weights b_j's to go with "j" we can write down, for example, a hypothesis related to $\mathcal{H}^*_{01}$ in the form

$$(2.3.1)\quad \mathcal{H}^{**}_{01}: \sum_i a_i p_{ijk} \big/ p_{ojk} = \text{an assumed function of } b_j \times \text{an unknown function of "k"} = (\lambda + \mu b_j) q_{**k} \text{ (say)},$$

$$\text{or } = (\lambda_k + \mu_k b_j) \text{ (say) (which is more general)},$$

where q_{**k} and λ_k and μ_k are supposed to be unknown functions of "k", and λ and μ are supposed to be unknown constants. It is easy to see that (2.3.1) is a fruitful special case of (2.2.1) when "j" is structured. We can test this in the spirit of testing for linearity of regression, or assuming this as a model, we can go ahead and test the hypothesis that $\mu = 0$ (say) or $\mu_k = 0$ (say).

A hypothesis naturally related to $\mathcal{H}^*_{02}$, for example, would seem to be

$$(2.3.2)\quad \mathcal{H}^{**}_{02}: \sum_i a_i p_{ijo} \big/ p_{ojo} = \text{an assumed function of } b_j = \lambda + \mu b_j \text{ (say), and}$$

$$\sum_i a_i p_{iok} \big/ p_{ook} \text{ is independent of "k".}$$

In this case again, treating λ and μ as unknown, we can test this hypothesis (in the spirit of testing for linearity of regression for the "normal" case) or, assuming this as a model, we can go ahead and test for, say, $\mu = 0$.

In this set-up a hypothesis naturally related to $\mathcal{H}^*_{04}$ and more general than (2.3.1) has not so far suggested itself to us. It may be observed that (2.3.1) itself might, in a sense, be regarded as being related to $\mathcal{H}^*_{04}$.

2.4 <u>The case where "i", "j" and "k" are all structured</u>.— As in section 2.3, if, for concreteness, we fix on "i" (say), then the hypothesis of independence of "i" with respect to "j" and "k" of the types (2.2.3), (2.2.5) and (2.2.7) (as also the tests) will stay the same as before. If these are tested and rejected then in subsequent experiments with the same type of material we have other important and interesting possibilities. Assuming a set of weights b_j's to go with "j" and c_k's with "k", we can write down, for example, a hypothesis related to $\mathcal{H}^*_{04}$ in the form

(2.4.1) $\mathcal{H}^{***}_{04}: \sum_i a_i p_{ijk} / p_{ojk} =$ an assumed function of b_j and $c_k = \lambda + \mu b_j + \nu c_k$ (a linear function, say, for the sake of concreteness and simplicity of illustration).

We can test this hypothesis (in the spirit of testing for linearity of regression in "normal" analysis of variance), in this case, treating λ, μ and ν as unknown and using a χ^2-test. Also assuming (2.4.1) as a model (on top of (2.1)) we can also test, for example, the hypothesis that $\mu = 0$ or that $\nu = 0$. Likewise, the one related to (2.2.5) would seem to be

$$(2.4.2)\quad \mathcal{H}_{02}^{***}: \sum_i a_i p_{ijo}\Big/ p_{ojo} = \text{an assumed function of } b_j = \lambda_1 + \mu b_j \quad (\text{say})$$

$$\text{and } \sum_i a_i p_{iok}\Big/ p_{ook} = \text{an assumed function of } c_k = \lambda_2 + \nu c_k \quad (\text{say}).$$

In this case also, treating λ_1, λ_2, μ and ν as unknown, we can test this hypothesis (in the spirit of testing for linearity of regression in the "normal" case) or, assuming this as a model we can go ahead and test for, say, $\mu = 0$ and/or $\nu = 0$.

Similarly, the one related to (2.2.3), although not too natural in this case, would seem to be

$$(2.4.3)\quad \mathcal{H}_{01}^{***}: \sum_i a_i p_{ijk}\big/ p_{ojk} = \text{an assumed function of } b_j \times \text{ an unknown function of "k"} = (\lambda + \mu b_j) q_{**k} \quad (\text{say})$$

$$\text{or } = \lambda_k + \mu_k b_j \quad (\text{say, to make it more general}),$$

where q_{**k}, λ_k and μ_k are supposed to be an unknown function of "k", and λ and μ are also supposed to be unknown constants. We can test this in the spirit of testing for linearity of regression, or, assuming this as a model we can go ahead and test for, say, $\mu = 0$ or $\mu_k = 0$. We can also set up a similar hypothesis by interchanging "j" and "k" and replacing b_j's by c_k's. It is easy to see that if both "j" and "k" are structured, then (2.4.1) and (2.4.2) are among the natural hypotheses to test. (2.4.3) is not too natural but is permissible. On the other hand, (2.4.3) would be quite natural if "k" were unstructured while "j" is structured. Likewise, a hypothesis similar to (2.4.3) in which the roles of "j" and "k" were interchanged and b_j's were replaced by c_k's would be natural if "j" were unstructured while "k" were structured. Nevertheless, even if "j" and "k" were both structured the hypothesis (2.4.3), one similar to this with "j" and "k" interchanged and with c_k in place of b_j's, and also one in which the right side is just replaced by $q_{**k} \times q_{*j*}$, would all be permissible, although none of them would be too natural.

The reader must have perceived that in the study of association in this nonparametric set-up we have been working in the spirit of "regression" (i.e., "prediction") rather than in the spirit of "correlation." In other words, we have not been trying to use a single measure for any of the various types of association. We feel that such a single measure has only a limited use and not too much use in the nonparametric set-up. However, such single measures (which come out as the noncentrality parameters in the asymptotic power functions of the respective tests for independence of various kinds) and their limited uses have already been discussed in [1,2,4,5].

3. A three-way table (ijk) in which "i" is a variate or response and "j" and "k" are factors. Let us consider a product-multinomial distribution

$$\Phi = \prod_{j,k} \left[\frac{n_{ojk}!}{\prod_i n_{ijk}!} \prod_i p_{ijk}^{n_{ijk}} \right], \tag{3.1}$$

where $\sum_i p_{ijk} = p_{ojk} = 1$ and $\sum_i n_{ijk} = n_{ojk}$ (fixed).

3.1 The case where "i","j" and "k" are all unstructured.--The hypothesis of no interaction between "j" and "k" means, essentially, that for a given "i", there is a lesser number of unknown parameters than would be given by all allowable (jk) combinations. Two specializations seem to be fruitful (and they are in the same spirit as in ordinary analysis of variance). From here on we must keep in mind that the unknown parameters or parametric functions introduced on the right sides must be such that (a) $0 < p_{ijk} < 1$ for all allowable (ijk) combinations and (b) $\sum_i p_{ijk} = 1$, again for all allowable (jk) combinations. The same kind of constraint also carries over into tables of dimensions higher than three, whenever we have factorial situations to deal with. The specializations are as follows.

$$\mathcal{H}_{06}: p_{ijk} = q_{i*k} q_{ij*} \tag{3.1.1}$$

and

$$\mathcal{H}^*_{06}: p_{ijk} = q^*_{i*k} + q^*_{ij*} . \tag{3.1.2}$$

The physical interpretations are as follows. For (3.1.1),

$p_{ijk}/p_{ij'k}$ is independent of "k" and $p_{ijk}/p_{ijk'}$ is independent of "j", and, for (3.1.2), $(p_{ijk} - p_{ij'k})$ is independent of "k" and $(p_{ijk} - p_{ijk'})$ is independent of "j".

Now, if it is a complete design, then summing both sides of (3.1.1) over "j" and "k" separately and then jointly it is easy to check that (3.1.1) can be rewritten (letting cb stand for complete block) in the equivalent form

$$(3.1.3) \qquad \mathcal{H}_{06}^{(cb)}: p_{ijk} = p_{iok}p_{ijo}/p_{ioo} ,$$

where it must be remembered that none of p_{iok}, p_{ojk}, or p_{ioo} is a probability, each being based on summation over the levels of a factor, i.e., a summation over subscripts belonging to different multinomial distributions. Thus (3.1.3) is only formally similar to (2.1.1), but it is easy to check that (3.1.3) would be completely identical with the condition for no partial association between "j" and "k", given "i", if "j" and "k" were variates and not factors as in the present case. Thus the no-interaction hypothesis in the form (3.1.3) (but only in the case of a complete design) has a tie-up with the no partial association hypothesis in case "j" and "k" were variates.

Next consider (3.1.2). Let $s^{(k)}$ = no. of "j"'s that go with a given "k", $t^{(j)}$ = no. of "k"'s that go with a given "j" and $u = \sum_{k=1}^{t} s^{(k)} = \sum_{j=1}^{s} t^{(j)}$ = no. of "jk" combinations. If it is a complete design then $s^{(k)} = s$, $t^{(j)} = t$, and $u = st$.

In this case, summing (3.1.2) over "j" and "k" separately and next over both "j" and "k" it is easy to check that (3.1.2) can be rewritten in the equivalent form

$$(3.1.4) \qquad \mathcal{H}_{06}^{(cb)*}: \; p_{ijk} = \frac{1}{s}\, p_{iok} + \frac{1}{t}\, p_{ijo} - \frac{1}{st}\, p_{ioo} \, .$$

Going back to $\mathcal{H}_{06}$ (expressed in the form (3.1.1)) we can test the narrower hypothesis that q_{ij*} is independent of "j",or, in other words, that p_{ijk} is independent of "j", which, without any loss of generality, we can write in either of the equivalent forms

$$(3.1.5) \qquad \mathcal{H}_{07}: \; p_{ijk} = q_{i*k}$$

or

$$(3.1.6) \qquad \mathcal{H}_{07}: \; p_{ijk} = \frac{1}{s^{(k)}}\, p_{iok} \, .$$

It is easy to check that, expressed in either of the equivalent forms (3.1.5) or (3.1.6), we shall eventually get the same hypothesis if we start from $\mathcal{H}_{06}^{*}$ (expressed in the form (3.1.2)) and try to state the narrower hypothesis that q_{ij*}^{*} is independent of "j", or, in other words, that p_{ijk} is independent of "j". Now, assuming for concreteness that "j" stands for treatments and "k" for blocks in a (possibly) incomplete block design (i) $\mathcal{H}_{06}$ (expressed in the form (3.1.1)) and (ii) $\mathcal{H}_{06}^{*}$ (again expressed in the form (3.1.2)) state respectively the hypothesis or model of no interaction (i) in the multiplicative sense and (ii) in the additive sense. $\mathcal{H}_{07}$, given by (3.1.6), states the hypothesis of no treatment effect. It is open to us (depending upon past knowledge) (a) to start from (2.1) as the model and test as a hypothesis either $\mathcal{H}_{06}$ or $\mathcal{H}^*_{06}$ or directly even $\mathcal{H}_{07}$, or (b) to start from a model which is (2.1) together with either $\mathcal{H}_{06}$ or $\mathcal{H}^*_{06}$, and then (in either case) to test $\mathcal{H}_{07}$ as a hypothesis.

3.2 <u>The case where "i" is structured but not necessarily "j" or "k"</u>.--In this case, the natural analogues of $\mathcal{H}_{06}$ and $\mathcal{H}^*_{06}$ (with a given set of a_i's) would seem to be

(3.2.1) $$\mathcal{H}_{08}: \sum_i a_i p_{ijk} = q_{**k} q_{*j*}$$

and

(3.2.2) $$\mathcal{H}^*_{08}: \sum_i a_i p_{ijk} = q^*_{**k} + q^*_{*j*} \ .$$

Remembering that q^*_{**k} and q^*_{*j*} are completely unknown functions of "j" and "k" respectively, we can rewrite (3.2.2) in the equivalent form

(3.2.3) $$\mathcal{H}^*_{08}: \sum_i a_i p_{ijk} = \sum_i a_i q^*_{i*k} + \sum_i a_i q^*_{ij*}$$

$$\text{or} \quad \sum_i a_i [p_{ijk} - (q^*_{i*k} + q^*_{ij*})] = 0 \ .$$

It is now easy to see that $\mathcal{H}^*_{06}$ of (3.1.2) $\Longrightarrow \mathcal{H}^*_{08}$ but not the other way around, but that (3.2.3) for (all sets of a_i's) $\Longrightarrow \mathcal{H}^*_{06}$. On the other hand, neither $\mathcal{H}_{06} \Longrightarrow \mathcal{H}_{08}$ nor the other way around (even for all sets of a_i's). Suppose we ask the question, what happens if (3.2.1) is to hold for all sets of a_i's? Rewriting the right side as, say, $q_{*j*} \sum_i a_i q_{i*k}$, we can rewrite (3.2.1) as

(3.2.4) $$\mathcal{H}_{08}: \sum_i a_i[p_{ijk} - q_{i*k}q_{*j*}] = 0 .$$

This means that if we set up a hypothesis

(3.2.5) $$\mathcal{H}_{081}: p_{ijk} = q_{i*k}q_{*j*}$$

then $\mathcal{H}_{081} \Longrightarrow \mathcal{H}_{08}$ but not the other way around. However, (3.2.4) (for all a_i's) $\Longrightarrow \mathcal{H}_{081}$ with a counterpart formed by interchanging "j" and "k" on the right side of (3.2.5). Thus it turns out that $\mathcal{H}_{08}$ does not have a natural tie-up with $\mathcal{H}_{06}$ in the sense in which $\mathcal{H}^*_{08}$ has a natural tie-up with $\mathcal{H}^*_{06}$. The tie-up of $\mathcal{H}_{08}$ (in this sense) is with $\mathcal{H}_{081}$ or its counterpart. However, if we sum up both sides of (3.2.5) over i, then since $p_{ojk} = 1$, we should have $1 = q_{*j*} \sum_i q_{i*k}$, whence $q_{*j*} = 1/\sum_i q_{i*k}$, which really means that both are pure constants (not depending on either "j" or "k" or "i"). Thus $\mathcal{H}_{081}$ is essentially $\mathcal{H}_{07}$, and its counterpart is essentially the counterpart of $\mathcal{H}_{07}$, and (3.2.1) (for all sets of a_i's) is the same as either (3.2.6) (for all sets of a_i's) or a statement (for all sets of a_i's) similar to (3.2.6) with q_{*j*} replacing q_{**k} on the right side.

Going back to $\mathcal{H}_{07}$ we write down its analogue in the form

(3.2.6) $$\mathcal{H}_{09}: \sum_i a_i p_{ijk} = q_{**k} ,$$

which, as in the case just preceding, can also be written in the equivalent form

(3.2.7) $$\mathcal{H}_{09}: \sum_i a_i[p_{ijk} - \frac{1}{s(k)} p_{iok}] = 0 \quad .$$

Again, as before, $\mathcal{H}_{07} \Longrightarrow \mathcal{H}_{09}$ but not the other way around. However, (3.2.7) (for all sets of a_i's) $\Longleftrightarrow$ (3.1.6). The other remarks made after (3.1.6) would also carry over to this case, covering (3.2.1), (3.2.2) and (3.2.6).

Going back to (3.2.1) we can set up as follows another hypothesis which has the same tie-up with $\mathcal{H}_{06}$ as $\mathcal{H}^*_{08}$ has with $\mathcal{H}^*_{06}$.

(3.2.8) $$\mathcal{H}_{082}: \prod_i p_{ijk}^{a_i} = q_{*j*}q_{**k} \quad .$$

Rewriting the right side as $\prod_i q_{ij*}^{a_i} q_{i*k}^{a_i}$ we can rewrite (3.2.8) as

(3.2.9) $$\mathcal{H}_{082}: \prod_i (p_{ijk}/q_{ij*}q_{i*k})^{a_i} = 1 \ .$$

It is easy to check that $\mathcal{H}_{06} \Longrightarrow \mathcal{H}_{082}$ but not the other way around, and also that (3.2.9) (for all a_i's) $\Longrightarrow \mathcal{H}_{06}$. It remains now to interpret the left side of (3.2.8). We notice that $\left[\prod_i p_{ijk}^{a_i}\right]^{\frac{1}{\sum_i a_i}}$ is a kind of weighted geometric mean of p_{ijk} (over "i") with a_i's for weights and that, since $\sum_i a_i$ is independent of "i", the left side of (3.2.8) is a certain power of this weighted geometric mean. We still don't know how meaningful or useful this interpretation of this model or hypothesis (3.2.8) would be in the end. We offer it quite tentatively at this stage.

3.3 <u>The case where "i" and "j" are both structured but not necessarily "k"</u>.--In this case (assuming a given set of weights b_j's to go with "j") the natural analogues of $\mathcal{H}_{08}$, $\mathcal{H}^*_{08}$ and $\mathcal{H}_{09}$ would seem to be

$$(3.3.1) \quad \mathcal{H}_{010}: \sum_i a_i p_{ijk} = q_{**k} \times \text{an assumed function of } b_j\text{'s}$$
$$= q_{**k} \times (\lambda + \mu b_j) \quad \text{(say)}$$
$$\text{or} \quad = \lambda_k + \mu_k b_j \text{ , say (which is more general),}$$

where λ_k and μ_k are supposed to be unknown functions of k and λ and μ just unknown constants,

$$(3.3.2) \qquad \mathcal{H}^*_{010}: \sum_i a_i p_{ijk} = q^*_{**k} + (\lambda + \mu b_j)$$

and

$$(3.3.3) \qquad \mathcal{H}_{011}: \sum_i a_i p_{ijk} = q_{**k} \quad .$$

The remarks on interpretation and inference procedures made after (3.1.6) would also carry over here. However, on the question of interpretation this case involves some additional possibilities (in terms of the conceptual universe or framework). While treating "k" as an unstructured factor it is open to us to treat "j" as a structured factor in a two-dimensional (jk) design or a concomitant variate in a one-way (k) classification. But for several reasons we shall, in this paper, skip this aspect and interpretation.

No matter which interpretation we make, if "k" is a factor and not a set of blocks (i.e., not a factor that is irrelevant and yet enters into the experiment in an unavoidable manner), then another meaningful hypothesis parallel to $\mathcal{H}_{011}$ would be

$$(3.3.4) \qquad \mathcal{H}_{012}: \sum_i a_i p_{ijk} = \lambda + \mu b_j \,,$$

(where λ and μ are, of course, assumed to be unknown). Going back to (3.3.1) we observe that in place of this it is open to us to consider the same kind of possibility that we did in the previous case in terms of $\mathcal{H}_{082}$ of (3.2.8). But we are not going to pursue it here. The same remark applies to later cases whenever we consider the multiplicative model or hypothesis of no interaction with responses that are structured and factors that are either all unstructured or some of them structured and the rest unstructured--for example, to (5.2.1), (5.2.3), (5.3.1) and (5.3.3).

3.4 The case where "i", "j" and "k" are all structured.--In this case, assuming furthermore a given set of weights c_k's to go with "k", it is possible to set up the same $\mathcal{H}_{010}$, $\mathcal{H}^*_{010}$, $\mathcal{H}_{011}$ and $\mathcal{H}_{012}$ and also similar ones in which the roles of "j" and "k" are interchanged and b_j's are replaced by c_k's. However, in this case, the more natural hypothesis would seem to be

$$(3.4.1) \qquad \mathcal{H}_{013}: \sum_i a_i p_{ijk} = \text{an assumed function of } b_j \text{ and } c_k = \lambda + \mu b_j + \nu c_k$$

(say, for the sake of concreteness and simplicity, where λ, μ and ν are, of course, supposed to be unknown). (i) Starting from (3.1) as a model we can test this hypothesis or directly one in which $\mu = 0$ and/or $\nu = 0$, or (ii) starting from (3.1) together with (3.4.1) as a model we can test the hypothesis that $\mu = 0$ and/or $\nu = 0$.

4. A four-way ($ijk\ell$) table in which "i", "j", "k" and "ℓ" are all responses or variates.

(4.1) $$\Phi = \frac{n!}{\prod_{i,j,k,\ell} n_{ijk\ell}!} \prod_{i,j,k,\ell} p_{ijk\ell}^{n_{ijk\ell}} ,$$

where $\sum_{i,j,k,\ell} p_{ijk\ell} = 1$ and $\sum_{i,j,k,\ell} n_{ijk\ell} = n$ (fixed).

Whether these variates are all unstructured, or all structured, or some of them are unstructured and the rest structured, there is much in common between this four-variate case and the three-variate set-up discussed in section 2. However, the four-variate set-up presents certain new features and we shall discuss a few of these, for purposes of illustration. These are by no means exhaustive. We shall consider, for example, for

4.1 the case where "i", "j", "k" and "ℓ" are all unstructured

(4.1.1) $$\mathcal{H}_{014}: \left.\begin{aligned} p_{ijko} &= p_{ioko}p_{ojko}/p_{ooko} \\ p_{ijo\ell} &= p_{ioo\ell}p_{ojo\ell}/p_{ooo\ell} \end{aligned}\right\} .$$

We shall also consider for

4.2 <u>the case where "ℓ" is unstructured and "i","j" and "k" are structured with given sets of weights</u> a_i's, b_j's <u>and</u> c_k's <u>to go with them</u>.

(4.2.1) $\mathcal{H}_{015}$: (i) $\sum_i a_i p_{ijko} / p_{ojko}$ = an assumed function of b_j and $c_k = \lambda + \mu b_j + \nu c_k$ (say)

or (ii) $\sum_j b_j p_{ijko} / p_{ioko} = \lambda^* + \mu^* a_i + \nu^* c_k$ (say)

together with (iii) $\sum_i a_i p_{ijo\ell} / p_{ojo\ell} = q_{***\ell} \times$ an assumed function of b_j

$= q_{***\ell} \times (\lambda^{**} + \mu^{**} b_j)$ (say)

or (iv) $\sum_j b_j p_{ijo\ell} / p_{ioo\ell} = q^*_{***\ell} \times (\lambda^{***} + \mu^{***} a_i)$ (say), where (i) is taken with (iii), or (ii) is taken with (iv). A more general possibility with λ_ℓ^{**} and μ_ℓ^{**} or with λ_ℓ^{***} and μ_ℓ^{***} could also be considered here, as in similar situations in previous cases.

5. <u>A five-way (ijkℓm) table in which "i" and "j" are responses or variates and "k","ℓ" and "m" are factors</u>.

$$(5.1) \qquad \Phi = \prod_{k,\ell,m} \left[\frac{n_{ook\ell m}!}{\prod_{i,j} n_{ijk\ell m}!} \prod_{i,j} p_{ijk\ell m}^{n_{ijk\ell m}} \right],$$

where $\sum_{i,j} p_{ijk\ell m} = p_{ook\ell m} = 1$ and $\sum_{i,j} n_{ijk\ell m} = n_{ook\ell m}$ (fixed).

Again, whether all these are structured or some of them structured and the rest unstructured, there is much in common between this set-up and the corresponding three-way set-up already discussed in section 3. However, there are some new features and we shall discuss a few of these as follows.

5.1 <u>The case where all are unstructured</u>.--We shall consider the hypothesis

$$\text{(5.1.1)} \quad \mathcal{H}_{016}: \left.\begin{aligned} p_{iok\ell m} &= q_{i**\ell m}\, q_{i*k*m}\, q_{i*k\ell*} \\ \text{and} \quad p_{ojk\ell m} &= q_{*j*\ell m}\, q_{*jk*m}\, q_{*jk\ell*} \end{aligned}\right\}$$

or

$$\text{(5.1.2)} \quad \mathcal{H}^*_{016}: \left.\begin{aligned} p_{iok\ell m} &= q^*_{i**\ell m} + q^*_{i*k*m} + q^*_{i*k\ell*} \\ \text{and} \quad p_{ijk\ell m} &= q^*_{*j*\ell m} + q^*_{*jk*m} + q^*_{*jk\ell*} \end{aligned}\right\},$$

and next, the hypothesis

$$\text{(5.1.3)} \quad \mathcal{H}_{017}: \left.\begin{aligned} p_{iok\ell m} &= q_{i*k**}\, q_{i**\ell*}\, q_{i***m} \\ \text{and} \quad p_{ojk\ell m} &= q_{*jk**}\, q_{*j*\ell*}\, q_{*j**m} \end{aligned}\right\}$$

or

$$\text{(5.1.4)} \quad \mathcal{H}^*_{017}: \left.\begin{aligned} p_{iok\ell m} &= q^*_{i*k**} + q^*_{i**\ell*} + q^*_{i***m} \\ \text{and} \quad p_{ojk\ell m} &= q^*_{*jk**} + q^*_{*j*\ell*} + q^*_{*j**m} \end{aligned}\right\}$$

and finally

(5.1.5) $\mathcal{H}_{018}$: the right side is independent of one or more of the factors "k","ℓ" and "m"

or

(5.1.6) $\mathcal{H}^*_{018}$: the right side is independent of one or more of the factors "k","ℓ" and "m."

It is easy to check that $\mathcal{H}_{016} \supset \mathcal{H}_{017} \supset \mathcal{H}_{018}$ and $\mathcal{H}^*_{016} \supset \mathcal{H}^*_{017} \supset \mathcal{H}^*_{018}$, and it seems to be natural to start from (5.1) and test $\mathcal{H}_{016}$ or $\mathcal{H}^*_{016}$ as a hypothesis, or from $\mathcal{H}_{016}$ or $\mathcal{H}^*_{016}$ as a model and test $\mathcal{H}_{017}$ or $\mathcal{H}^*_{017}$ as a hypothesis, or from $\mathcal{H}_{017}$ or $\mathcal{H}^*_{017}$ as a model and test $\mathcal{H}_{018}$ or $\mathcal{H}^*_{018}$ as a hypothesis. There are various interesting intermediate cases which we omit and which the reader can supply himself.

5.2 The case where "i" and "j" are structured and "k", "ℓ" and "m" are unstructured.--The analogues of the hypotheses in the previous case would, in this case, seem to be as follows (and in the same order).

$$(5.2.1)\quad \mathcal{H}_{019}: \left.\begin{aligned} \sum_i a_i p_{iok\ell m} &= q^{(1)}_{***\ell m} q^{(1)}_{**k*m} q^{(1)}_{**k\ell *} \\ \text{and} \quad \sum_j b_j p_{ojk\ell m} &= q^{(2)}_{***\ell m} q^{(2)}_{**k*m} q^{(2)}_{**k\ell *} \end{aligned}\right\}$$

or

$$(5.2.2)\quad \mathcal{H}^*_{019}: \left.\begin{aligned} \sum_i a_i p_{iok\ell m} &= q^{*(1)}_{***\ell m} + q^{*(1)}_{**k*m} + q^{*(1)}_{**k\ell *} \\ \text{and} \quad \sum_j b_j p_{ojk\ell m} &= q^{*(2)}_{***\ell m} + q^{*(2)}_{**k*m} + q^{*(2)}_{**k\ell *} \end{aligned}\right\} ,$$

$$(5.2.3)\quad \mathcal{H}_{020}: \left.\begin{aligned} \sum_i a_i p_{iok\ell m} &= q^{(1)}_{**k**} q^{(1)}_{***\ell *} q^{(1)}_{****m} \\ \text{and} \quad \sum_j b_j p_{ojk\ell m} &= q^{(2)}_{**k**} q^{(2)}_{***\ell *} q^{(2)}_{****m} \end{aligned}\right\}$$

or

$$(5.2.4)\quad \mathcal{H}^*_{o20}: \left.\begin{array}{l} \sum_i a_i p_{iok\ell m} = q^{*(1)}_{**k**} + q^{*(1)}_{***\ell *} + q^{*(1)}_{****m} \\ \text{and } \sum_j b_j p_{ojk\ell m} = q^{*(2)}_{**k**} + q^{*(2)}_{***\ell *} + q^{*(2)}_{****m} \end{array}\right\},$$

and finally

(5.2.5) $\mathcal{H}_{o21}$: the right side is independent of one or more of the factors "k","ℓ" and "m"

or

(5.2.6) $\mathcal{H}^*_{o21}$: the right side is independent of one or more of the factors "k","ℓ" and "m".

The remarks made after (5.1.6) would also apply here.

We shall consider a final

5.3 <u>case where "i","j",and "k" are structured while "ℓ" and "m" are unstructured</u>.--The natural analogues of the hypotheses of the previous case would, in this case, seem to be (in the proper order) as follows:

$$(5.3.1)\quad \mathcal{H}_{o22}: \left.\begin{array}{l} \sum_i a_i p_{iok\ell m} = (\lambda^{(1)}_{\ell m} + \mu^{(1)}_{\ell m}\ c_k) \quad \text{(say)} \\ \text{and } \sum_j b_j p_{ojk\ell m} = (\lambda^{(2)}_{\ell m} + \mu^{(2)}_{\ell m}\ c_k) \quad \text{(say)} \end{array}\right\}.$$

or

$$(5.3.2)\quad \mathcal{H}^*_{o22}: \left.\begin{array}{l} \sum_i a_i p_{iok\ell m} = q^{*(1)}_{***\ell m} + \lambda^{*(1)} + \mu^{*(1)} c_k \text{ (say)} \\ \text{and } \sum_j b_j p_{ojk\ell m} = q^{*(2)}_{***\ell m} + \lambda^{*(2)} + \mu^{*(2)} c_k \text{ (say)} \end{array}\right\},$$

$$(5.3.3)\quad \mathcal{H}_{023}: \sum_i a_i p_{iok\ell m} = (\lambda^{(1)}_{\ell *}\lambda^{(1)}_{*m} + \mu^{(1)}_{\ell *}\mu^{(1)}_{*m}\; c_k) \quad \text{and} \quad \sum_j b_j p_{ojk\ell m} = (\lambda^{(2)}_{\ell *}\lambda^{(2)}_{*m} + \mu^{(2)}_{\ell *}\mu^{(2)}_{*m}\; c_k)$$

or

$$(5.3.4)\quad \mathcal{H}^*_{023}: \sum_i a_i p_{iok\ell m} = q^{*(1)}_{***\ell *} + q^{*(1)}_{****m} + \lambda^{*(1)} + \mu^{*(1)}\; c_k \text{ (say)} \quad \text{and} \quad \sum_j b_j p_{ojk\ell m} = q^{*(2)}_{***\ell *} + q^{*(2)}_{****m} + \lambda^{*(2)} + \mu^{*(2)}\; c_k \text{ (say)},$$

and finally

(5.3.5) $\mathcal{H}_{024}$: the right side is independent of one or more of "ℓ" and "m" with or without the μ's being zero

or

(5.3.6) $\mathcal{H}^*_{024}$: the right side is independent of one or more of "ℓ" and "m" with or without the μ^*'s being zero.

The remarks made after (5.1.6) would also apply here as in the previous case. Going back to (5.3.1) it may be observed that it might also be regarded as a rather remote analogue of (5.2.11) given by

$$(5.2.11)\quad \mathcal{H}_{0191}: \sum_i a_i p_{iok\ell m} = q^{(1)}_{***\ell m}\, q^{(1)}_{**k**} \quad \text{and} \quad \sum_j b_j p_{ojk\ell m} = q^{(2)}_{***\ell m}\, q^{(2)}_{**k**},$$

(which is a special case of (5.2.1)) and likewise, (5.3.2) might also be regarded as a rather remote analogue of (5.2.21) given by

$$(5.2.21)\quad \mathcal{H}^{*}_{0191}: \sum_i a_i p_{iok\ell m} = q^{*(1)}_{***\ell m} + q^{*(1)}_{**k**}$$

$$\text{and} \quad \sum_j b_j p_{ojk\ell m} = q^{*(2)}_{***\ell m} + q^{*(2)}_{**k**} \quad ,$$

(which again is a special case of (5.2.2)). There is a question whether in this case there is anything more general than (5.3.1) or (5.3.2) but along the same lines. So far we have failed to hit on any such thing.

6. Concluding remarks. The foregoing discussion suggests the proper treatment for a complex experimental situation in which some of the responses or variates are structured, say v_1 in number, the rest unstructured, say v_2 in number, some of the factors are structured, say f_1 in number, while the rest are unstructured, say f_2 in number. In theory or principle the tests in terms of the appropriate χ^2-statistic (or difference of two χ^2's) can be set up for all the specific hypotheses mentioned, as also for the corresponding hypotheses in the larger framework indicated in this section and the asymptotic power functions can also be calculated [1,2, 4,5]. The actual form of the χ^2 has been given for only some of the specific hypotheses discussed in sections 2-5, and the asymptotic power functions for even fewer of such tests. More of both have been obtained later on and the derivation of the rest is under way. These will be presented in a subsequent publication.

The treatment given above is all in terms of single or product-multinomial distributions which, again, are based on sampling with replacement. That is no doubt somewhat unrealistic, the more realistic operation being sampling without replacement, which would generate a hypergeometric or product-hypergeometric distribution. However, even under sampling without replacement, the multinomial or product-multinomial distribution would be a good enough approximation if the sampling fraction were small. If the sampling fraction were not small, a development in terms of hypergeometric or product-hypergeometric distributions parallel to the above seems to be called for, and this is under consideration.

References

[1] E.L. Diamond, "Asymptotic power and independence of certain classes of tests on categorical data," Institute of Statistics, University of North Carolina, Mimeograph Series No. 196.

[2] S.K. Mitra, "Statistical analysis of categorical data," Institute of Statistics, University of North Carolina, Mimeograph Series No. 142.

[3] S.N. Roy and Marvin A. Kastenbaum, "On the hypothesis of 'no-interaction' in a multiway contingency table," Annals of Mathematical Statistics, vol. 27 (1956), pp. 749-757.

[4] S.N. Roy and S.K. Mitra, "An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis," Biometrika, col. 43 (1956), pp. 361-376.

[5] S.N. Roy, "Some aspects of multivariate analysis," John Wiley and Sons, New York (1958), Chapter 15.