ARTIKLER

75

ARTIKLER FRA STATISTISK SENTRALBYRÅ NR. 75

MULTIPLE COMPARISONS BY BINARY AND MULTINARY OBSERVATIONS

By Erling Sverdrup

MULTIPLE SAMMENLIKNINGER VED BINÆRE OG MULTINÆRE OBSERVASJONER

OSLO 1975

ISBN 82-537-0482-8

PREFACE

The Central Bureau of Statistics must continuously judge whether seeming tendencies express underlying realities. In this Article an attempt is made to develop a systematic method for treatment of such questions. It aims at situations which are very common in the practice of the Bureau, where the observations are binary or multinary. The method is similar to corresponding methods developed by Henry Scheffé for "linear-normal" situations, where each observation is a measurement.

Central Bureau of Statistics, Oslo, 15 May 1975

Petter Jakob Bjerve

FORORD

Statistisk Sentralbyrå må ofte vurdere hvorvidt tilsynelatende tendenser i et datamateriale gir uttrykk for underliggende realiteter. I denne artikkelen er det gjort et forsøk på å utvikle en systematisk metode for behandling av slike spørsmål. Den tar sikte på situasjoner som vanligvis opptrer ved Byråets undersøkelser, nemlig binære og multinære situasjoner. Metoden er analog til tilsvarende metoder utviklet av Henry Scheffé for "lineær-normale" situasjoner hvor enkeltobservasjonene er målinger.

Statistisk Sentralbyrå, Oslo, 15. mai 1975

Petter Jakob Bjerve

CONTENTS

Page

INNHOLD

1. Description of the multiple comparison rule

A. We shall be concerned with statistical data where each observation may take one of several exclusive forms. It is reasonable to call such observations "multinary", if there are only two forms they may be called "binary". The following examples will illustrate our point. In an employment investigation one is interested in observing, for each employed person, if they remain employed after one year and for each unemployed if they get employed within a year. Thus the forms are "employed after one year", "unemployed after one year". - In a general mortality investigation individuals are observed to die within or survive after one year. - If the investigation is directed toward death from tuberculosis each observation takes three forms, "dying within a year from tuberculosis", "dying within a year from other causes", "surviving the year". - For the purpose of studying relationships between allergy and type of mental ailment, each mental patient is observed to be allergic or not and to have one of three major types of mental ailments (neurotic, psychopatic, psychogenic). Hence each observation takes one of six different forms. - By investigation of a certain insecticide, the death or survival of insects exposed to the insecticide are observed.

In general, each observation can assume one of r exclusive forms $A_1, A_2, \dots A_r$, which are assumed to have probabilities $p_1, p_2, \dots p_r$, $\sum_1^r p_j = 1$.

The observations can, based on a priori considerations, be classified in s different groups, in each of which the forms $A_1, A_2, \dots$ are given, and which are homogeneous with respect to the probabilities $p_1, p_2, \dots$. Thus one shall suppose that in group a, $a=1,2, \dots s$, the observations can take r_a forms

(1) $\qquad A_{a1}, \dots, A_{ar_a}$

with probabilities

(2) $\qquad p_{a1}, \dots, p_{ar_a}$

We assume all the multinary observations to be independent. We shall include the case where there is just one group (s=1). That may be the case in the example of observations of mental patients mentioned above, in which case $r_a=r=6$.

In the above examples of mortality investigations there may be 10 five-year age groups from age 20 to 70. Hence s=10 and all $r_a=2$.

In the example of the employment investigation the data may be classified in 8 five-year groups from age 20 to age 60. The purpose of the investigation may be to observe the change in tendency to remain employed or to obtain employment over 10 years. Thus at the beginning of each of the 10 years we have employed and unemployed in each age group and we observe if they are employed at the end of the year. There are s=8x2x10=160 groups and all $r_a=2$. (We shall use this example below and then find it convenient to use our method separately for each of the 8x2 groups, still keeping full control of the probability of making an error. Then we have s=10.)

In the example with insecticide the insects may under 3 different environmental conditions be exposed to doses of 5 different concentrations. Hence s=15 and $r_a=2$.

B. In the set-up defined above there are certain types of comparisons we are interested in making.

In the employment example let us confine ourself to one age group and consider those persons unemployed at the beginning of each of the 10 years. Then p_{a1}, p_{a2} are the probabilities that a person unemployed at the beginning of the year a shall obtain, resp. not obtain, employment during the year. Let us write $p_{a1}=p_a$, for short, and $p_{a2}=1-p_a$. Then we might be interested in changes in the chances p_a of obtaining employment. We are interested in investigating if $p_{a+1}-p_a > 0$ or $p_{a+1}-p_a < 0$; a=1,2 ..., s. We are of course only interested in making such statements to the extent to which they are justified by the limited information available from our data. Then, however, the situation may arise, where we are not willing to say that there has been a change over each of the (say) 4 one-year periods from 1970-1974, but that we are willing to state that it has been a definite decrease in the chance of obtaining employment from (say) 1970 to 1973. Thus we are interested in all comparisons

$p_b - p_a$ where $a=1,2,\ldots b-1$, $b=2,3,\ldots,10$. Of course, we are interested in making the probability of a false statement small. Hence our statistical method ought to be such that the probability for any a and b of stating that $p_b > p_a$ when indeed $p_b < p_a$, should be small, say at most 0.05. (Note that we consider the probability of the union of all the (45) statements just pronounced.) Now, for the sake of illustration we shall stretch this example further than is perhaps realistic in this special case. We suggest then, that there might be an interest in whether the increased chance of obtaining employment is accelerating ("escalated"), $p_{a+1}-2p_a+p_{a-1} > 0$, or about to retard, $p_{a+1}-2p_a+p_{a-1} < 0$. We might also be interested in the relationship between the chance p_a of getting a job and some other quantity W varying over the 10 years. (And it will not worry us that perhaps we have "discovered" the possibility of such a relationship after having looked at the data, I do not think, however, that will be the case in such typical social applications as the present example represents. After all, the economists are aware of so many relationships.) Suppose that W takes the values W_1, W_2, ..., W_{10} over the 10 years of observation. Then we may be interested in whether there is positive or negative covariation between p and W, measured, say, by the covariance $\sum_{a=1}^{10} p_a(W_a-\bar{W})$ ($\bar{W}$ = mean of W_a). - Summing up, we are interested in discovering "<u>contrasts</u>" f among the p_a, i.e. relations $f(p) = \sum_{a=1}^{10} f_a p_a > 0$, where $\sum_{a=1}^{10} f_a = 0$. (Note that $\Sigma f_a = 0$ in all the relations mentioned above, $1-1 = 0$, $1-2+1 = 0$, $\Sigma(W_a-\bar{W}) = 0$.) We want to construct a method which is such that <u>the probability is at most (say) 0.05 of stating</u> $f(p) = \Sigma f_a p_a > 0$ <u>for at least one</u> (f_1, ..., f_{10}) <u>for which</u> $\Sigma f_a p_a \leq 0$. We shall give a formal general definition of a contrast below.

Consider now, the above example of mental patients. Let p_{i1}, p_{i2} be the probabilities that a specific patient has mental ailment i and does not suffer, resp. suffers, from allergy. Thus $p_{i1}+p_{i2} = p_i$ is the probability of suffering from i; whereas, $q_1 = \sum_i p_{i1}$, $q_2 = \Sigma\, p_{i2}$ are the probabilities of not suffering, resp. suffering from allergy. Now, if mental ailment and allergy where independent events, then $p_{ij} = p_i \cdot q_j$. Thus $p_{i2} > p_i q_2$ means that mental patients of type i are apt to have allergy. We <u>are interested in finding positive ties between type of ailments and allergy type</u>. Thus a typical contrast f is the following type of function of all p_{ij}, $f(p) = p_{ij} - \sum_k p_{ik} \sum_\ell p_{\ell j}$.

C. Refering to the general set-up in A with independent multinary observations, we shall now define the class of "contrasts" we are interested in. Note that in the above example of employment investigation all the interesting contrasts we were interested in were zero if there was no change over time, i.e. p_a = constant. In the example of mental patients the contrasts were zero in case of independence. We shall call such specifications of our model "null-states". It does not mean that we have any a priori belief in those special states, perhaps we even know for certain that they cannot be true.

In the general set up we shall define the null-state by the following restrictions on the parameters p_{aj},

(3) $p_{aj} = \phi_{aj}(\theta_1, \ldots, \theta_t),\ j=1,2,\ldots, r_a,\ a=1,2,\ldots,s.$

We shall include the case where the null state has the form $p_{aj} = \phi_{aj}$ with ϕ_{aj} fixed numbers. Then we set t =0. For $t > 0$ we assume that $\theta = (\theta_1, \ldots, \theta_t)$ can take any value in an open set Θ in the t-dimensional space. - Now, any function f(p) of $p = (p_{11}, \ldots, p_{sr_s})$ which is such that if we insert $p = \phi = (\phi_{11}, \ldots, \phi_{sr_s})$ then

(4) $f(\phi(\theta)) = 0$

identically in θ, will be called a contrast relatively to the null state.

Thus, in the employment example we may let the scalar θ be the constant probability of being employed within a year, hence

$$p_a = p_{a1} = \theta$$

$$1-p_a = p_{a2} = 1-\theta$$

We have t=1 and we see that any form $\Sigma f_a p_a$ with $\Sigma f_a = 0$ is a contrast since $\Sigma f_a \theta = \theta \Sigma f_a = 0$.

In the example with mental patients we let θ_1, θ_2, $1-\theta_1-\theta_2$ be the probabilities of having ailments 1, 2, 3 respectively and θ', $1-\theta'$ the probabilities of not having, resp. having allergy. Thus ϕ is defined by

$$p_{i1} = \theta_i\,\theta', \quad p_{i2} = \theta_i(1-\theta'), \quad i=1,2,$$

$$p_{31} = (1-\theta_1-\theta_2)\theta', \quad p_{32} = (1-\theta_1-\theta_2)\,(1-\theta').$$

Thus t=3 in this case and we easily see that any $p_{ij}-p_i \cdot q_j$ is a contrast.

Let us also consider the problem of what we can read out of a mortality table for a certain period. The table gives the probability q_x for a person x years of age of dying within a year, for

$x = x_0, \ldots, x_0+s-1$. Let L_x be the number of persons reaching x years of age in the period and let D_x be the number og those who die before age x+1. Then q_x is estimated by $\hat{q}_x = D_x/L_x$. Suppose we are interested in studying the change in q_x with age. Then the null state is that $q_x = \theta$. Any q_b-q_a is a contrast, and s=s, $r_a=2$, t=1. Suppose, however, that from age x_0 to x_0+s-1 we were pretty sure that q_x increases with age. We are interested in how it changes, then we may put $q_x = \theta_1+\theta_2 x$ and $(q_{a+1}-q_a) - (q_{b+1}-q_b)$, $q_{a+1}-2q_a+q_a$ are contrasts, but q_b-q_a is not.

Returning to the general set-up we shall be interested in linear contrasts f, where f has the form

$$(5) \qquad f(p) = \sum_{a=1}^{s} \sum_{j=1}^{r_a} f_{aj}p_{aj}$$

with the f_{aj} as constants, independent of p. We shall, however, also be interested in smooth contrasts where f has continuous first order derivatives.

D. The multiple comparison method we shall propose has two versions.

Version 1. (null state estimated variance). Pick out any special contrast in which you have become interested, perhaps after having looked at the data. Treat it by the classical normal approximation theory (with one exception, see below), i.e. replace the probabilities p_{aj} in the contrast by the relative frequencies from our data. We then obtain an "estimated contrast". Find the variance of this estimated contrast. This will again depend on the probabilities p_{aj}. Replace again these p_{aj} by the maximum likelihood estimates $\hat{\phi}_{aj}$ of p_{aj} in the nullstate. The square root of this quantity is the "estimated standard deviation". Declare now the contrast to be positive if the estimated contrast is $\geq \sqrt{z}$ multiplied by the estimated standard deviation; where z is $1-\varepsilon$ fractile of the chi-square distribution with

$$R-s-t, \quad (R = \sum_{1}^{s} r_a)$$

degrees of freedom and ε is the level of significance. Repeat the same for all the other interesting contrasts. Thus the one important distinction from the classical normal approximation theory is that we should use $\sqrt{z}$ defined above in place of the $1-\varepsilon$ fractile of the normal distribution. Then it will be shown below that asymptotically the probability of making a false statement among all the statements we might possibly make is at most ε.

Let us take an example to illustrate our method. In the employment example above we were interested in a contrast of the type $p_b - p_a$. The estimated contrast is $p_b^* - p_a^*$, where p_a^* (p_b^*) is the relative number of unemployed at the beginning of the year a (b) which obtain employment during the year. Now the variance of the estimated contrast is

$$\text{var} \quad (p_b^* - p_a^*) = \frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}$$

where n_a (n_b) is the number of unemployed at the beginning of the year a (b). Let now N_a be those out of the n_a unemployed which obtain employment during the year a. Then $\hat{\phi} = \sum_1^{10} N_j / \sum_1^{10} n_j$ would be the maximum likelihood estimate of any p_a, $a=1,2,\ldots,10$ in the null state (when all p_a are assumed equal). Hence the estimated variance is

$$\hat{\sigma}^2 = \text{est.var} \ (p_b^* - p_a^*) = \left(\frac{1}{n_a} + \frac{1}{n_b}\right)\hat{\phi}(1-\hat{\phi})$$

We have R=20, s=10, t=1. Hence the number of degrees of freedom is R-s-t=9. With $\varepsilon=0.05$ we get z=16.9. Thus p_b^* should be declared $> p_a^*$ if $p_b^* - p_a^* > 4.1 \cdot \hat{\sigma}$. (In the normal approximation theory we would have used 2 in stead of 4.1). In general we declare that $\sum_1^{10} c_j p_j > 0$ if

$$\Sigma c_j p_j^* > 4.1 \cdot \sqrt{\hat{\phi}(1-\hat{\phi}) \ \Sigma c_j^2 / n_j}$$

Let us return to the general method. It is noteworthy that R-s-t is the <u>number of degrees of freedom you would have used if you would test the null state as a "null hypothesis" by the chi-square-goodness-of-fit test</u>. Now it will be shown below that there exists no significant linear contrast by the rule above unless we get significance by the chi-square-goodness-of-fit test. (This is a strict algebraic relationship. There are no approximations or probabilities involved.) The chi-square statistic in question is given below in section 2.E equation (14) (where $\hat{\phi}_{aj} = \phi_{aj}(\hat{\theta})$ is given by equation (7) in section 2.C).

Since it is sometimes rather time consuming to look around for significant contrasts (there may be none), it is therefore occasionally time saving to <u>start by performing a chi-square-goodness-of-fit test of the null-state</u>. If significance is not obtained, we can just <u>declare the data uninteresting</u>.

Version 2. We proceed as by version 1, but after having found the expression for the variance of the contrast, the p_{aj} in this expression is replaced by relative frequencies from our data, instead of $\hat{\phi}_{aj}$.

Also in this case there exists no significant contrast unless we get significance by the chi-square-goodness-of-fit test, but now the modified chi-square statistic defined by equation (15) in section 2.E must be usual. ($\hat{\phi}_{aj}$ in this equation equals $\phi_{aj}(\hat{\theta})$ where $\hat{\theta}$ is given by equation (8) in section 2.C.)

E. We shall go into details with some specific examples, using version 1 of our rule.

Example 1. Seasonal variation in birth rate. We shall study births by month and to simplify we shall confine ourself to one year. Live born children in Norway in 1970 had the following distribution on months.

Jan.	Febr.	March	April	May	June	July	Aug.
5 384	4 977	5 866	6 150	5 617	5 404	5 321	5 279

Sept.	Oct.	Nov.	Dec.	Sum
5 390	4 957	4 837	5 369	64 551

We denote these numbers by $N_1, \ldots, N_{12}$, $n = \Sigma N_i$.

We are interested in knowing if there are any months with high "fertility" or low fertility and any significant changes of fertility over months. This very manner of formulating the problem suggests using uniform distribution over month as null state. It is the only reason for being interested in such a null state, we are not interested in testing it as a "null hypothesis". Let p_j be the probability that a child born in 1970 shall be born in month j. Our null state is

$$p_1 = \ldots = p_{12} = \frac{1}{12}$$

and hence t=0. We order the months according to decreasing frequencies of births,

Month	Births	Differences		
April	6 150			
March	5 866	284	533	
May	5 617	249	462	
June	5 404	213		
September	5 390	14		
January	5 384	6	338	
December	5 369	15		640
July	5 321	48		
August	5 279	42		
February	4 977	302		
October	4 957	20		
November	4 837	120		
	64 551			

We are above all interested in comparing any pair of two months, hence $f = p_i - p_j$. We have $p_i^* = N_i/n$, R-s-t = 11, $\sqrt{z} = 4.4$ with $\varepsilon = 0.05$ and

$$\sigma_f^2 = \text{var } (p_i^* - p_j^*) = \frac{1}{n} (2 \frac{1}{12} \cdot \frac{11}{12} + 2 \frac{1}{12} \cdot \frac{1}{12}) = \frac{1}{n} \frac{1}{6}$$

We find $\sqrt{z}\, \hat{\sigma}_f = \frac{1}{\sqrt{n}} 1.796$. It is convenient to operate with absolute numbers. Hence $n\sigma_f^2 = \text{var } (N_i - N_j) = n/6, \sqrt{zn}\, \hat{\sigma}_f = 456$. Comparing with the table above, it is seen that we should be willing state that April is more fertile than all other months except March. March is more fertile than the other 10 months except May. Between the 10 other months there are no clear distinction except that May is more fertile than February, October and November.

Having looked at the data it is reasonable to ask if the period March-April-May is more fertile than the other months. This amounts to asking if the probability of beeing born in one of these three months is $> 3/12 = \frac{1}{4}$. The estimated contrast is then given by

$$N\hat{f} = N_3 + N_4 + N_5 - N/4 = 17633 - 16138 = 1495, \quad \sigma_f^2 = N \cdot \frac{1}{4} \cdot \frac{3}{4} = 12103, \quad \hat{\sigma}_f = 110,$$

$\sqrt{z}\, \hat{\sigma}_f = 484$. Since $1495 > 484$ we can state that these three months are more fertile than the others. (We might, of course, have pooled the months of March and April with any other month than May and obtained a similar result, but these results would not be interesting.)

Example 2. We have observed the following frequencies of live births (indigeneous) in Oslo.

Month	1943	1944	1945	1946	1943-46
January	282	339	433	355	1 409
February	253	351	330	414	1 348
March	271	356	319	529	1 475
April	299	372	439	539	1 649
May	313	341	342	463	1 459
June	260	368	351	425	1 404
July	290	404	319	408	1 421
August	277	322	362	371	1 332
September	284	351	331	408	1 374
October	341	346	265	429	1 381
November	362	335	324	352	1 373
December	297	264	391	358	1 310
Jan.-Dec.	3 529	4 149	4 206	5 051	16 935

Table 1

			Percentages		
Month	1943	1944	1945	1946	1943-46
January	7.99	8.17	10.29	7.03	8.32
February	7.17	8.46	7.85	8.20	7.96
March	7.68	8.58	7.58	10.47	8.71
April	8.47	8.97	10.44	10.67	9.74
May	8.87	8.28	8.13	9.17	8.61
June	7.37	8.87	8.35	8.41	8.29
July	8.22	9.74	7.58	8.08	8.39
August	7.85	7.76	8.61	7.35	7.87
September	8.05	8.46	7.87	8.08	8.11
October	9.66	8.34	6.30	8.49	8.15
November	10.26	8.07	7.70	6.97	8.11
December	8.41	6.36	9.30	7.09	7.74
Jan.-Dec.	100.00	100.00	100.00	100.00	100.00

Table 2

We are interested in "irregular" changes in births apart from a possible regular seasonal variation and the trend of increased number of births from one year to another.

Having looked at the data, we become specially interested in the figur from March 1946 (table 2), which is far above what seems to have been normal for the month of March.

Let p_{aj} be the probability that a birth in the year a takes place in the month j; $j = 1, 2, \ldots, 12$; $a = 1, 2, 3, 4$. N_{aj} is the actual number of births in (a, j) (Table 1). Obviously we are interested in contrasts relatively to the nullstate

$$p_{aj} = \phi_{aj} = \theta_j; \quad j = 1, 2, \ldots, 11$$

$$\phi_{a12} = 1-\theta_1- \ldots -\theta_{11}$$

Let us write $\phi_{aj} = \phi_j$ and introduce

$$n_a = \sum_j N_{aj}; \quad N_j = \sum_a N_{aj}; \quad N = \Sigma n_a = \Sigma N_i$$

The $p^*_{aj} = N_{aj}/n_a$ are given in table 2. $\hat{\phi}_j = N_j/N$ is the maximum likelihood in the nullstate and is given in the last column in table 2.

We are interested in a contrast of the form $f = p_{aj}-p_{bj}$. We have

$$\sigma_f^2 = \mathrm{var}\ (p^*_{aj}-p^*_{bj}) = \frac{p_{aj}(1-p_{aj})}{n_a} + \frac{p_{bj}(1-p_{bj})}{n_b}$$

and

$$\hat{\sigma}_f^2 = (\frac{1}{n_a} + \frac{1}{n_b})\ \hat{\phi}_j\ (1-\hat{\phi}_j)$$

The number of degrees of freedom is R-s-t = 48-4-11 = 33 = (3x11) and the 0.95 fractile of the chi-square distribution with 33 degrees of freedom is $z = 47.4$, $\sqrt{z} = 6.88$. With a=4 (1946), b=1 (1943), j=3, we get $\hat{\sigma}_f = 0.0062$, $\sqrt{z}\ \hat{\sigma}_f = 0.0426$, whereas the difference $p^*_{43}-p^*_{13} = 0.0282$. The difference is not significant. It should be emphasized that even if we had stopped here and did not investigate any other contrasts, we should use $\sqrt{z} = 6.88$ and not 1.96 (the level 5% critical value in the normal distribution), since we looked at the data before we decided to be interested in just the month of March. (With 1.96 instead of $\sqrt{z}$, we would have got $1.96\hat{\sigma}_f = 0.0121 < 0.0282$, hence significance.)

However, since we use $\sqrt{z}$, we are free to use "any means" to make 10.49 significant. It is then natural to look at

$$f(p^*) = p^*_{43}- \sum_1^3 p^*_{a3}C_a\ , \sum_1^3 C_a = 1$$

Hence

$$\hat{\sigma}_f^2 = (\frac{1}{n_4} + \sum_1^3 C_a^2/n_a)\ \hat{\phi}_3\ (1-\hat{\phi}_3)$$

and hopefully minimize this with respect to C_1, C_2, C_3, giving $C_a = n_a/m$, $m = \sum_1^3 n_a$ and

$$\hat{\sigma}_f^2 = (\frac{1}{n_4} + \frac{1}{m})\ \hat{\phi}_3\ (1-\hat{\phi}_3), \quad \sum_1^3 p^*_{a3}C_a = \sum_1^3 N_{a3}/m$$

Hence $f(p^*) = 0.0251$, $\hat{\sigma}_f = 0.0473$, $\sqrt{z}\ \hat{\sigma}_f = 3.25$. Still there is no significance. We have to be satisfied with that. Could we have obtained significance at all? Perhaps we should have asked that question at once. We get for the chi-square-goodness-of-fit

$$Z = \sum_{a,j} (N_{aj}-n_a\hat{\phi}_j)^2/n_a\hat{\phi}_j = N\left[\sum_a \frac{1}{n_a} \sum_j \frac{N_{aj}^2}{N_j} - 1\right] = 192.1 > 46.6$$

Hence some less interesting significant contrasts must be hidden in the material. It is important, however, that anybody who finds something interesting is free to test it, provided he uses $\sqrt{z} = 6.88$ and not 1.96 as fractile.

Example 3. We return to the employment investigation mentioned in section A above. Employed and unemployed (2 groups) in 8 five-year age-groups from age 20-60 at the beginning of each of 10 years are observed through the year to see of they remain (obtain) employment. Thus there are 2x8x10 = 160 groups and the observations are binary. Let $p_{\alpha\tau}$ (resp. $p'_{\alpha\tau}$) be the probability of an employed (resp. unemployed) person in age-group α at the beginning of the year τ to be employed a year later; $\alpha = 1, 2, \ldots, 8$; $\tau = 1, 2, \ldots, 10$. Since we are interested in the change over time, it is natural to let the null state be

$$p_{\alpha\tau} = \theta_\alpha, \quad p_{\alpha\tau} = \theta'_\alpha$$

Thus there are t=16 parameters, R=320 cells and s=160 distributions, R-s-t = 144. Following our general rule, we have R-s-t = 144 and with level $\varepsilon=0.05$ that z=173, $\sqrt{z}=13.2$. Let now $n_{\alpha\tau}$ ($n'_{\alpha\tau}$) be the number of employed (unemployed) in age-group α at the beginning of the year τ and $N_{\alpha\tau}$ ($N'_{\alpha\tau}$) number of those who are employed at the end of the year. Then $p^*_{\alpha\tau} = N_{\alpha\tau}/n_{\alpha\tau}$, $p'^*_{\alpha\tau} = N'_{\alpha\tau}/n'_{\alpha\tau}$ are the a priori estimates of $p_{\alpha\tau}$ and $p'_{\alpha\tau}$, whereas $\hat{\phi}_\alpha = \sum_\tau N_{\alpha\tau}/\Sigma\ n_{\alpha\tau}$ and $\hat{\phi}'_{\alpha\tau} = \Sigma\ N'_{\alpha\tau}/\Sigma\ n'_{\alpha\tau}$ are our nullstate estimates. According to our rule we should declare $\sum_{\alpha,\tau} p_{\alpha\tau}f_{\alpha\tau} + \sum_{\alpha,\tau} p'_{\alpha\tau}f'_{\alpha\tau} > 0$, where $\sum_\tau f_{\alpha\tau} = \sum_\tau f'_{\alpha\tau} = 0$ for all α, for which

$$\Sigma p^*_{\alpha\tau} f_{\alpha\tau} + \Sigma p'^*_{\alpha\tau} f_{\alpha\tau} > 13.2\sqrt{\sum_\alpha \hat{\phi}_\alpha (1-\hat{\phi}_\alpha) \sum_\tau f_{\alpha\tau}^2/n_{\alpha\tau} + \sum_\alpha \hat{\phi}'_{\alpha\tau}(1-\hat{\phi}'_{\alpha\tau}) \sum_\tau f'^2_{\alpha\tau}/n'_{\alpha\tau}}$$

This would make it possible to study not only the variation of $p_{\alpha\tau}$ and $p'_{\alpha\tau}$ over time for each age group α, but also to compare the changes for different age groups.

If we are not interested in comparisons of changes for different age groups, then we could proceed as follows. Treat each of 16 groups according to age and state of employment separately and use level $1-(1-\varepsilon)^{1/16} = 0.0031$. Use a $\sqrt{z}$ corresponding to this level with 9 degrees of freedom. That gives $\sqrt{z} = 5.0$. Thus we shall declare $\sum_\tau p_{\alpha\tau} f_{\alpha\tau} > 0$ if

$$\sum_\tau p^*_{\alpha\tau} f_{\alpha\tau} > 5\sqrt{\hat{\phi}_\alpha (1-\hat{\phi}_\alpha) \sum_\tau f_{\alpha\tau}^2/n_{\alpha\tau}}$$

(and similarly with $p'_{\alpha\tau}$). The corresponding chi-square test is

$$Z_\alpha = \sum_\tau \frac{(N_{\alpha\tau} - n_{\alpha\tau}\hat{\phi}_\alpha)^2}{n_{\alpha\tau}\hat{\phi}_\alpha(1-\hat{\phi}_\alpha)} > z = 24.8$$

(and $Z'_\alpha > 24.8$ where Z'_α defined as Z_α but with N', n', ϕ' instead of N, n, ϕ).

It is clear that both methods suggested would have a level ε, but that the last method would be more sensitive to contrasts of special interest.

F. The statistical method proposed can be said to suffer from two shortcomings.

(i). In particular in social applications it will usually be natural to relate the multinary variables considered to some other variables and thus construct some kind of econometric model where the multinary variables are involved. In principle one should then be able to construct a more sensitive method, i.e. a method which would reveal more contrasts then with our present method. Often the work in constructing such an econometric model and a statistical rule based on it, is formidable. In the present authors opinion the method proposed above could therefore be useful. Suppose the method does not select a certain contrast as significant, then an economist is warned. If he wants to declare it significant anyhow, he must do so with expressed reference to some other variables or circumstances which are not taken into account.

(ii). Another objection to the method proposed is that it safeguards against false statements not only about interesting contrasts, but also some uninteresting contrasts. Having defined a null state, you have to drag along with all contrasts relatively to that nullstate. This again would also impair sensitivity. This is really a problem inherent in many statistical methods. It is difficult to gauge both the model and the decision space to fit the special situation you are interested in. There are, however, some flexibility in our method, as demonstrated by example 3 above. We obtained higher sensitivity by kicking out some uninteresting contrasts a priori.

2. Properties of the general multiple comparison rule

In 1953 Henry Scheffe [1] proposed a method of judging all contrasts in a linear-normal situation in such a manner that the probability of making a false statement is kept under control. It is noteworthy that this can be shown to be true for any values of the parameters, not only under the null-hypothesis (see also [2]).

The present paper extends this result to the multinomial situation.

A. The result of n independent multinary trials are observed. The series of trials may be divided into s sequences such that there are n_a trials in the i-th sequence; $a = 1, 2, \ldots, s$; $\Sigma n_a = n$. Each of the trials in the i-th sequence may result in one of r_a mutually exclusive events

$$A_{a1}, \ldots, A_{ar_a}$$

with propabilities

$$p_{a1}, \ldots, p_{ar_a}; \quad \sum_{j=1}^{r_a} p_{aj} = 1 \tag{1}$$

We assume a priori that all p_{aj} are between 0 and 1. The observed number of times the r_i events occur are

$$N_{a1}, \ldots, N_{ar_a}, \quad \sum_{j=1}^{r_a} N_{aj} = n_a, \tag{2}$$

respectively. Let $R = \sum_{a=1}^{s} r_a$.

B. The null-state H is to effect that the p_{aj} are specified functions

(3) $$p_{aj} = \phi_{aj}(\theta); \quad \theta \epsilon \Theta; \quad j = 1, 2, \ldots, r_a; \quad a = 1, 2, \ldots, s$$

of a parameter $\theta = (\theta_1, \ldots, \theta_t)$, where θ varies in an open set Θ in the t-space. We assume that the ϕ_{aj} have continuous second order derivatives, and that the rank of the Rxt matrix

(4) $$\left\{ \frac{\partial \phi_{aj}(\theta)}{\partial \theta_i} \right\} \quad (a,j) = (1,1), \ldots, (s, r_s), \; i = 1, \ldots, t,$$

is t.

A function $f(p)$ of $p = \{p_{11}, \ldots, p_{sr_s}\}$ is a contrast relatively to H if $f(\phi) = 0$ for all θ. It will be called smooth if it has continuous first order derivatives

(5) $$f_{aj} = \frac{\partial f}{\partial p_{aj}}$$

We shall consider a class $\mathcal{F}$ of smooth contrasts. Two cases will be treated. Case (i). $\mathcal{F}$ is the set of all (or some) linear contrasts

(6) $$f = \Sigma \, f_{aj} p_{aj}$$

Thus in this case the f_{aj} are independent of p. Case (ii). $\mathcal{F}$ is such that the class of all f_{aj} obtained by varying f in $\mathcal{F}$ is equicontinous.

C. The statistical method can be described as follows. First the maximum likelihood estimates $\hat{\theta}$ under H are found as solutions of

(7) $$\sum_{a,j} \frac{N_{aj}}{\phi_{aj}(\theta)} \cdot \frac{\partial \phi_{aj}(\hat{\theta})}{\partial \theta_i} = 0; \quad i = 1, 2, \ldots, t$$

Alternatively $\hat{\theta}$ is a modified minimum chi-square estimator, given by

(8) $$\sum_{a,j} \frac{N_{aj} - n_a \phi_{aj}(\hat{\theta})}{N_{aj}} \, n_a \, \frac{\partial \phi_{aj}(\hat{\theta})}{\partial \theta_i} = 0; \quad i = 1, 2, \ldots, t,$$

We do not care whether $\hat{\theta}$ actually maximizes the likelihood

$$L = \prod \phi_{aj}(\theta)^{N_{aj}}$$

or, alternatively, minimizes,

$$\chi^2 = \Sigma \frac{(N_{aj}-n_a\phi_{aj}(\theta))^2}{N_{aj}}$$

We assume that for all $\{N_{aj}\}$, (7) (or (8)) has either one or no solution. We shall let all n_a go to infinity in such a manner that $n_a/n = g_a > 0$. We assume that the probability that (7) (or (8)) has one solution goes to 1. When (7) (or (8)) has no solution we can let $\hat{\theta}$ have any value (e.g. such that it actually maximizes L). It can then be proved that plim $\hat{\theta}=\theta$. (That is what we need. It is rather trivial that something much stronger can be proved.)

Let $p^*_{aj} = N_{aj}/n_a$ and $\hat{\phi}_{aj} = \phi_{aj}(\hat{\theta})$. We find in case (i),

(9) $$\sigma^2_f(p) = \text{var } f(p^*) = \sum_a n_a\left[\sum_j f^2_{aj}p_{aj} - (\Sigma f_{aj}p_{aj})^2\right]$$

In case (ii), $\sigma^2_f(p)$ can be found by linearizing $f(p^*)$ with respect to p^*-p. Now define

(10) $$\hat{\sigma}_f = \sigma_f(\hat{\theta}), \quad \sigma^*_f = \sigma_f(p^*)$$

These two quantities will be called respectively the <u>null state estimated</u> and a <u>priori estimated standard deviation</u>.

The rule consists in stating that $f(p) > 0$ for all those $f \in \mathcal{F}$ for which

(11) $$f(p^*) > \sqrt{z}\,\hat{\sigma}_f$$

where z is the $1-\varepsilon$ fractile of the chi-square distribution with R-s-t degrees of freedom. Alternatively we may use σ^*_f on the right hand side of (11).

It should be noted that if we want to test $f(p) \leq 0$ with a specified form f selected in advance, then we would have used the $1-\varepsilon$ fractile (or perhaps $1-\frac{\varepsilon}{2}$ fractile) for the normal distribution instead of $\sqrt{z}$ (one degree of freedom for z).

<u>D</u>. We shall prove below in case (i) and (ii) <u>that in the limit the probability of making one false statement; i.e. stating that f > 0 for an</u> <u>$f \in \mathcal{F}$ for</u> which $f \leq 0$; is asymptotically $\leq \varepsilon$. More precisely

(12) $$\limsup_{n\to\infty} \Pr\Big(\bigcup_{f(p)\leq 0} (f(p^*) > \sqrt{z}\,\sigma^*_f)\Big) \leq \varepsilon$$

if $p = p^{(n)}$ approaches some $p^{(o)}$ as $n \to \infty$. Of course this includes the particular case when p is kept constant. The reason for stating the more general result is that it is sometimes desirable to let p go to some ϕ as n goes to infinity (e.g. in such a manner that all $\sqrt{n}\,(p_{aj}-\phi_{aj})$ are kept constant).

We shall prove that (12) still holds with $\hat{\sigma}_f$ instead of σ_f^*, <u>but now we must let $p^{(n)}$ approach some ϕ.</u>

It will also be seen that if the p_{aj} are kept constant such that $f(p) \neq 0$ for all $f \in \mathcal{F}$, then the left hand side of (12) is 0. (in case (i) this can only take place if $\mathcal{F}$ is some subset of all linear contrasts).

If in particular $p = \phi$ then all statements $f(p) > 0$ are wrong and in case (i) we shall prove that the probability of making such a statement approaches ε, i.e.

$$(13) \qquad \lim \Pr\Big(\bigcup_f \big(f(p^*) > \sqrt{z}\,\hat{\sigma}_f\big)\Big) = \varepsilon$$

(13) is also true with $\hat{\sigma}_f$ replaced by σ_f^*.

<u>E.</u> In case (i) with $\mathcal{F}$ consisting of all linear contrasts, it will be proved that <u>if the null state estimated variance is used then some contrast will be declared positive if and only if</u>

$$(14) \qquad Z = \sum_{a,j} \frac{(N_{aj}-n_a\hat{\phi}_{aj})^2}{n_a\hat{\phi}_{aj}} > z$$

<u>where $\hat{\phi}_{aj} = \phi_{aj}(\hat{\theta})$ is a maximum likelihood estimate. If an a priori estimated variance is used then a contrast will be declared positive if and only if</u>

$$(15) \qquad Z = \sum_{a,j} \frac{(N_{aj}-n_a\hat{\phi}_{aj})^2}{N_{aj}} > z$$

<u>where $\hat{\phi}_{aj} = \phi_{aj}(\hat{\theta})$ is a modified minimum chi-square estimate.</u>

Note that these relations between the multiple comparison rules on the one hand side and (14) and (15) on the other hand side are <u>purely algebraic</u>. They are strictly true, there are no approximations involved and it is not a probability statement.

In case (ii) asymptotically (11) takes place for some $f \in \mathcal{F}$ only if (14) (or (15)) is true.

This suggests that both in case (i) and case (ii) one might first check if (14) (or (15)) is true and only if such is the case go on to apply (11). <u>Thus the test proposed is a refinement of the classical Karl Pearsson's significance test.</u>

Note that if the a priori estimated variance is used, then the estimate $\hat{\theta}$ is not needed in connection with the multiple comparison rule, It is only needed for checking (15).

It is of interest to consider the special case of homogeneity testing. Then $r_1 = \ldots = r_s = r$ and we choose as a null state that $p_{a1}, \ldots, p_{ar}$ are independent of a. This can be written

$$\phi_{aj} = \theta_j; \quad j = 1, 2, \ldots, r-1, \quad \phi_{ar} = 1-\theta_1 - \ldots - \theta_{r-1} = \theta_r$$

We then get from (7) the maximum likelihood estimates

$$\hat{\phi}_{aj} = \hat{\theta}_j = \sum_a N_{aj}/n = N_j/n \tag{16}$$

and from (8) the minimum modified chi-square estimates

$$\hat{\phi}_{aj} = \hat{\theta}_j = \bar{p}_j/\sum_1^r \bar{p}_i \tag{17}$$

where $\bar{p}_j$ is the harmonic mean of the $p^*_{aj} = N_{aj}/n_a$; $a = 1, 2, \ldots, s$.

$$\bar{p}_j = n/\sum_a \frac{n_a}{p^*_{aj}} \tag{18}$$

The chi-square statistics are respectively for null-state estimated and a priori estimated variances,

$$Z = n\left(\sum \frac{N_{aj}^2}{n_a N_s} - 1\right), \quad Z = n(-1+1/(\Sigma\bar{p}_j)^2) \tag{19}$$

They are found from (14) and (15) respectively. Now it is seen that $\Sigma p_{aj} f_{aj}$ is a contrast if and only if $\sum_a f_{aj} = 0$ for $j = 1, 2, \ldots, r$. According to the general rule with a priori estimated variances this should be declared > 0 if

$$\Sigma\, p^*_{aj} f_{aj} > \sqrt{z}\sqrt{\sum_a (\Sigma f^2_{aj} p^*_{aj} - (\Sigma f_{aj} p^*_{aj})^2)} \tag{20}$$

where z is determined with (r-1) (s-1) degrees of freedom. (20) will take place for some f_{aj} if and only if

$$(\Sigma\bar{p}_j)^2 < \left(1 + \frac{z}{n}\right)^{-1} \tag{21}$$

where $\bar{p}_j$ is given by (18).

This special case has been treated by Goodman [4]. Reiersöl [3] has treated a situation from probit-analysis with multiple comparison, which is, however, not completely covered by the present set-up.

3. Proof of the assertions about the general rule

A. In sections A-F we shall treat the case (i) when $\mathcal{F}$ consists of linear contrasts and null-state estimates are used. We introduce

$$(1) \qquad Y_{aj} = \frac{N_{aj} - n_a \hat{\phi}_{aj}}{\sqrt{n_a \hat{\phi}_{aj}}}$$

and it is well known that in the limit, when $n \to \infty$; with $n_a = ng_a$, $g_a > 0$; then $Z = \sum_{a,j} Y_{aj}^2$ has a chi-square distribution with R-s-t degrees of freedom. We now have in case (i) (see 2.B), since $f(\hat{\phi}) = 0$,

$$f(p^*) = \Sigma f_{aj} \left(\frac{N_{aj}}{n_a} - \hat{\phi}_{aj}\right)$$

We introduce $h_{aj} = f_{aj} \sqrt{\frac{\hat{\phi}_{aj}}{g_a}}$ and get

$$(2) \qquad f(p^*) = \frac{1}{\sqrt{n}} \Sigma h_{aj} Y_{aj}$$

B. Let us now consider,

$$(3) \qquad \sigma_f^2(p) = \mathrm{var}\, f(p^*) = \sum_1^s \frac{1}{n_a} \mathrm{var} \sum_1^{r_a} f_{aj} N_{aj} / \sqrt{n_a}$$

We have

$$(4) \qquad \mathrm{cov}\,(N_{aj}/\sqrt{n_a},\ N_{ak}/\sqrt{n_a}) = \begin{cases} p_{aj}(1-p_{aj}) & \text{if } h = j, \\ -\, p_{aj} p_{ak} & \text{if } k \neq j. \end{cases}$$

We then get for $\hat{\sigma}_f^2 = \hat{\sigma}_f^2(\hat{\phi})$

$$(5) \qquad \hat{\sigma}_f^2 = \frac{1}{n} \sum_{a=1}^{s} \sum_{j,h} (\delta_{jk} - \sqrt{\hat{\phi}_{aj}}\sqrt{\hat{\phi}_{ak}}) h_{aj} h_{ak}$$

(where we have made use of the Krönecker δ)

C. Below we shall, in order to facilitate the introduction of matrix notations, replace (a, j) by a single letter i, such that $i = 1, 2, \ldots, R$ represents (a, j) in lexical ordering. Hence $N_{aj} = N_i$, $p_{aj} = p_i$, $\phi_{aj}(\theta) = \phi_i(\theta)$, $f_{aj} = f_i$, $h_{aj} = h_i$. We write also n_i and g_i in place of n_a and g_a. Thus n_i and g_i are constants on sections of length $r_1, r_2 \ldots, r_s$, respectively. We denote the sections by $S_1, \ldots, S_a$ respectively, and have

$$\sum_{i \in S_a} N_i = n_a, \quad \sum_{i \in S_a} p_i = 1$$

We can now write (1)

$$(6) \qquad Y_i = \frac{N_i - n_i \hat{\phi}_i}{\sqrt{n_i \hat{\phi}_i}}; \quad i = 1, 2, \ldots, R$$

Now, let b denote a matrix of order Rxs, the a-th column of which is $(0, \ldots 0, \sqrt{\hat{\phi}_{a1}}, \ldots, \sqrt{\hat{\phi}_{ar_a}}, 0 \ldots 0)$ (The column starts with $\sum_1^{a-1} r_i$ zeros). We then get from (2)

$$(7) \qquad f(p^*) = \frac{1}{\sqrt{n}} \sum_1^R h_i Y_i = \frac{1}{\sqrt{n}} h'Y$$

and from (5)

$$(8) \qquad \hat{\sigma}_f^2 = \frac{1}{n} h'(I - bb')h$$

D. From the contrast property of f we have,

$$(9) \qquad \sum f_i \phi_i = 0,$$

hence

$$(10) \qquad \sum_1^R f_i \frac{\partial \phi_i(\hat{\theta})}{\partial \hat{\theta}_j} = 0; \quad j = 1, 2, \ldots, t$$

We introduce the matrix

$$(11) \qquad B_{ij} = \sqrt{\frac{g_i}{\hat{\phi}_i}} \frac{\partial \phi_i(\hat{\theta})}{\partial \hat{\theta}_j}; \quad i = 1, 2, \ldots, R; \quad j = 1, 2, \ldots, t$$

It is seen that $B = \{B_{ij}\}$ is the matrix $\frac{\partial \phi_i(\hat{\theta})}{\partial \hat{\theta}_j}$ multiplied by a diagonal non-singular matrix. Hence, by 2.B, B must have rank t. We can write (10)

(12) $h'B = 0$

From

(13) $\sum_{i \in S_a} \phi_i(\hat{\theta}) = 1; \quad a = 1, 2, \ldots, s,$

we get by derivation with respect to $\hat{\theta}_j$; $j = 1, 2, \ldots, t,$

(14) $b'B = 0$

E. Since B has full rank, the space V_t spanned by the columns of B is a t-dimensional subspace of the R-dimensional vector space V_R. Let H be a Rxt matrix such that its columns constitute an orthonormal basis for V_t. Then of course $H'H = I$ and since by (12) and (14) h and all columns of b are perpendicular to V_t we have

$h'H = 0$, (15); $b'H = 0$ (16).

From (16) it is seen that the columns of the matrix (H, b) has orthogonal columns. We complete it and obtain an orthogonal matrix

(17) $K = (G, H, b)$

of order RxR.

Let us now introduce

(18) $d = K'h, \quad V = K'Y$

Then we have from (7)

(19) $\sqrt{n}\, f(p^*) = h'Y = d'V$

(15) reduces to

$$0 = h'H = d'K'H = (0, \ldots, 0, d_{R-s-t+1}, \ldots d_{R-s}, 0 \ldots 0)$$

Hence,

(20) $\quad d_{R-s-t+1} = \dots = d_{R-s} = 0$

From equation (8) we get $n\hat{\sigma}_f^2 = h'h - h'bb'h = d'd - d'K'bb'Kd$. But

$$K'b = \begin{Bmatrix} G' \\ H' \\ b' \end{Bmatrix} b = \begin{Bmatrix} 0 \\ 0 \\ b'b \end{Bmatrix},$$

which combined with $b'b = I$ and (20) gives

(21) $$n\hat{\sigma}_f^2 = \sum_1^{R-s-t} d_i^2$$

For V given by (18) we have

$$V = \begin{Bmatrix} G' \\ H' \\ b' \end{Bmatrix} Y$$

But by (6), the a-th component of $b'Y$ is

$$\sum_{i \in S_a} (N_i - n_i\hat{\phi}_i)/\sqrt{n_i} \, ,$$

which from (13) and since n_i is constant, equals 0. Thus $V_{R-s+1} = \dots = V_R = 0$ and by (19)

(22) $$\sqrt{n}\, f(p^*) = \sum_1^{R-s-t} d_i V_i$$

By (22) and (21) the criterion 2.(11) for stating that $f(p) > 0$ reduces to

(23) $$\sum_1^{R-s-t} d_i V_i > \sqrt{z \sum_{i=1}^{R-s-t} d_i^2}$$

Thus we make no statement if and only if

$$\sum_1^{R-s-t} d_i V_i \leq \sqrt{z \sum_1^{R-s-t} d_i^2}$$

for all d. But for given Σd_i^2, the maximum of the left hand side is, by Schwartz inequality,

$$\sqrt{\sum_1^{R-s-t} d_i^2 \sum_1^{R-s-t} V_i^2}$$

Thus we make no statement if and only if

$$(24) \qquad \sum_1^{R-s-t} V_i^2 \leq z$$

Now we make use of the fact that $\hat{\theta}$ is a maximum likelihood estimate in the null state, i.e. satisfies 2.(7), which can be written

$$(25) \qquad \sum_{i=1}^{R} \frac{N_i}{\hat{\phi}_i} \frac{\partial \phi_i(\hat{\theta})}{\partial \hat{\theta}_j} = 0; \quad j = 1, 2, \ldots, t$$

By derivation of (13) with respect to $\hat{\theta}_j$, multiplying by n_a, summing over all a, and subtracting from (25), we get

$$(26) \qquad B'Y = 0$$

Hence $H'Y = 0$ and by (18), $V_{R-s-t+1} = \ldots = V_{R-s} = 0$. Thus (24) is the same as

$$(27) \qquad Z = \sum_1^{R} Y_i^2 \leq z$$

Hence we have proved the statement in 2.E in case (i).

F. Since in the null state Z is chi-square distributed with R-s-t degrees of freedom, we have proved the assertion in 2.D about the probability of making a false statement in case (i) when $p = \phi$ and null state estimated variance is used.

G. We shall still consider case (i), but we now assume that we use a priori estimated variances in the multiple comparison rule. The derivation in A-F can then be repeated with the following changes.

h_{aj} is now defined $= f_{aj}\sqrt{\frac{p^*_{aj}}{g_a}}$ and (1) is replaced by

$$(1)' \qquad Y_{aj} = \frac{N_{aj} - n_a \hat{\phi}_{aj}}{\sqrt{N_{aj}}}$$

with the corresponding change in (6). In the definition of b, $\hat{\phi}_{aj}$ is replaced by p^*_{aj}. The definition of B_{ij} in (11) is replaced by

$$(11)' \qquad B_{ij} = \sqrt{\frac{g_i}{p_i^*}}\,\frac{\partial\phi_i(\hat\theta)}{\partial\hat\theta_j}$$

From 2.(8) we get (26) with Y defined by (1)'. Hence we get the statement in 2.E and 3.F in case (i) with a priori estimated variance replacing null state estimated variance.

H. Now let $p \neq \phi$ in case (i). We consider the multiple comparison rule when a priori estimated variance is used. Let

$$(28) \qquad X_i = \frac{N_i - n_i p_i}{\sqrt{N_i}}$$

Then

$$(29) \qquad \sqrt{n}\, f(p^*) = \sqrt{n} \sum_1^R f_i \frac{N_i}{n_i} = \sum_1^R h_i X_i + \sqrt{n}\, f(p)$$

where $h_i = f_i \sqrt{\frac{p_i^*}{g_i}}$. Hence the probability of a false statement can be written,

$$(30) \qquad \Pr\Big(\bigcup_{f(p)\le 0} \{\sum_1^R h_i X_i + \sqrt{n}\, f(p) \ge \sqrt{zn}\, \sigma_f^*\}\Big)$$

where the union is taken over all $f \in \mathcal{F}$ such that $f(p) \le 0$, for given p. Obviously for fixed $p \neq \phi$ such that $f(p) \neq 0$ for all $f \in \mathcal{F}$, (30) goes to 0.

Replace now p by $p^{(n)}$ and let $p^{(n)} \to p$. Then it can be shown that $\text{plim}\, p^* = p$. Then from (8),

$$(31) \qquad \text{plim}\, n\sigma_f^{*2} = \bar h(I - \bar b \bar b')\bar h = \sigma_f^2(p) \quad \text{(say)}$$

where $\bar h_i = f_i \sqrt{\frac{p_i}{g_i}}$ and $\bar b$ is b with p^* replaced by p. Now it is seen that (30) is

$$\le \Pr\Big(\bigcup_{f(p)\le 0} \{\bar h' X \ge \sqrt{zn}\, \sigma_f^*\}\Big) \le$$

$$(32) \qquad \le \Pr\Big(\bigcup_f \{h' X \ge \sqrt{zn}\, \sigma_f^*\}\Big),$$

where in the last expression the union is taken over all linear contrasts f. From (30), (31), (32) we get

$$(33) \qquad \text{limsup}\, \Pr\,(\text{false statement}) \le \Pr\Big(\bigcup_f \{\bar h' X \ge \sqrt{z}\, \sigma_f(p)\}\Big)$$

where X now denotes a multinormally distributed variable with mean zero and covariance matrix $(I-\bar{b}\bar{b}')$. Below we shall drop the bars over h and b, and we shall define B by

$$(34) \qquad B_{ij} = \sqrt{\frac{g_i}{p_i}}\ \frac{\partial\phi_i(\Theta)}{\partial\theta_j}$$

instead of by (11) or (11)'. The columns of the matrix H is now an orthogonal basis for the column space of this new B. Then the relations (12), (13), (15), (16) still hold. G is such that $K = (G, H, b)$ is an RxR orthogonal matrix and we define $d = Kh$, $W = K'X$. Then $h'X = d'W$ and the covariance matrix of W equals $K'(I-bb')K$, which has zero elements except for the R-s first elements of the main diagonal which equal 1. Hence $W_{R-s+1} = \ldots = W_{R-s} = 0$ with probability 1. From $h'H = 0$ we get $d_{R-s-t+1} = \ldots = d_{R-s} = 0$. The right hand side of (33) then reduces to

$$(35) \qquad \Pr(\bigcup_d \{\sum_1^{R-s-t} d_j W_j > \sqrt{z}\sqrt{\sum_1^{R-s-t} d_i^2}\})$$

where $W_1, \ldots, W_{R-s-t}$ are independent normal (0, 1). As in section E (eq. (23)-(24)), (35) reduces to

$$(36) \qquad \Pr(\sum_1^{R-s-t} W_j^2 > z) = \varepsilon$$

This proves an assertion in 2.D in case (i). The assertion in 2.D concerning the multiple comparison rule with null state estimated variances is proved in the same manner.

<u>I</u>. From the fact that the right hand side of (33) equals ε and from (29) we get

$$(37) \qquad \lim \Pr(\bigcup_f \{\sqrt{n}\ f(p^*) - \sqrt{n}\ f(p) \geq \sqrt{z}\ \sigma_f^*\}) = \varepsilon$$

<u>which gives us a simultaneous confidence interval for all contrasts</u>.

<u>J</u>. We have above not gone into details about the arguments involving limits in probability and limits in distribution. We shall be even more superficial below when using these kinds of argument and we shall defer a more rigorous treatment to a later publication.

We turn to case (ii) and assume that $p \to \phi$ as $n \to \infty$. We denote the partial derivatives of f(p) with respect to p_i by $f_i(p)$. We can now go through the derivation as above and note the following alterations.

Equation (7) is asymptotically true since by the Taylor expansion.

(38) $$\sqrt{n}\ f(p^{*}) = \sqrt{n}\ \Sigma f_i(p')\ (\frac{N_i}{n_i} - \hat{\phi}) = \Sigma f_i(p') \sqrt{\frac{\hat{\phi}_i}{g_i}}\ Y_i$$

(where p' is "between" p^{*} and $\hat{\phi}$). Thus we can let

(39) $$h_i = f_i(\hat{\phi}) \sqrt{\frac{\hat{\phi}}{g_i}}$$

in (7).

From

(40) $$\sqrt{n}\ f(p^{*}) = \sqrt{n}\ f(p) + \sum_{1}^{R} f_i(p'')(p_i^{*}-p)$$

we see that $\sqrt{n}\ f(p^{*})$ has asymptotic variance given by (8).

By derivation of the identity $f(\phi(\theta)) = 0$ we get that (10) is rigorously true with $f_i = f_i(\hat{\phi})$. The equicontinuity assumption i 2.B leads to (23) being true in the limit in probability uniformly with respect to d, and this is needed to obtain that (24) is true in the limit in probability.

From (40) it is also seen that (29) is true asymptotically.

REFERENCES

[1] Henry Scheffe: A method of judging all contrasts in the analysis of variance. Biometrika (1953) vol. 40, p. 87-104.

[2] Erling Sverdrup: Multiple decision theory. Lecture Note Series n. 15, Matematisk Institut, Aarhus 1969.

[3] Olav Reiersöl: Linear and non-linear multiple comparison in logit analysis. Biometrika 48. p. 359-365 (1961).

[4] Leo A. Goodman: Simultaneous confidence intervals for contrasts among multinomial population. Ann. of math. Stat. vol. 35 (1964) p. 716-725.

Utkommet i serien ART

Issued in the series Artikler fra Statistisk Sentralbyrå (ART)

Nr. 65 Jan Byfuglien: Bosettingskart over Norge 1970: Grunnlag, innhold og bruk *Map of the Population Distribution of Norway 1970:Basis, Contents and Use* 1974 43 s. kr. 7,00

" 66 John Dagsvik: Etterhåndsstratifisering og estimering innen delbestander *Post-Stratification and Estimation within Subpopulations* 1974 49 s. kr. 7,00

" 67 Arne Rideng: Klassifisering av kommunene i Norge 1974 *Classification of the Municipalities of Norway* 56 s. kr. 7,00

" 68 Erik Biørn: Estimating the Flexibility of the Marginal Utility og Money: An Errors-in-Variables Approach *Estimering av pengenes grensenyttefleksibilitet: et opplegg med feil i de variable* 1974 18 s. kr. 5,00

" 70 Inger Gabrielsen: Aktuelle skattetall 1974 *Current Tax Data* 1974 73 s. kr. 8,00

" 71 Vidar Ringstad: Some Empirical Evidence on the Decreasing Scale Elasticity *Noen resultater for produktfunksjoner med fallende passuskoeffisient for norsk bergverk og industri* 1974 20 s. kr. 5,00

" 72 Jon D. Engebretsen: En modell for analyse av utviklingen i de direkte skatter: Skattemodellen i MODIS IV *A Model for Analysis of the Development in Direct Taxes: Tax Model in MODIS IV* 1974 65 s. kr. 8,00

" 73 Lars Østby: Hvem flytter i Norge? Tendenser i flyttergruppens sammensetning etter 1950 *The Migrants in Norway Trends in the Composition of the Migrant Group after 1950* 1975 23 s. kr. 5,00

" 74 Erik Biørn: Avskrivningsregler og prisen på bruk av realkapital *Depreciation Rules and the User Cost of Capital* 1975 46 s. kr. 7,00

" 75 Erling Sverdrup: Multiple Comparisons by Binary and Multinary Observations *Multiple sammenlikninger ved binære og multinære observasjoner* 1975 33 s. kr. 7,00

Fullstendig oversikt over tidligere nummer av serien Artikler finnes i nr. 70.