TESTING EQUALITY OF PROPORTION OF SUCCESS OF SEVERAL CORRELATED BINOMIAL VARIATES

by

Peter A. Lachenbruch and Jan Perry

Department of Biostatistics and
Hand Rehabilitation Center

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 767

August 1971

TESTING EQUALITY OF PROPORTION OF SUCCESS OF SEVERAL CORRELATED BINOMIAL VARIATES

by

Peter A. Lachenbruch[1] and Jan Perry[2]

An important problem often arises when a series of correlated binomial variates are observed and it is desired to test if they have the same probability of success. A simple solution is proposed and an example is given.

Consider the following problem. In physical therapy, it is thought that the sweat and sensation in the hand are linked together. If sweat is present, then sensation will be also, and vice versa. If either is missing, the other should be also. In a study to examine this phenomenon, 35 patients at the Hand Rehabilitation Center at UNC were studied. These patients all had severed nerves in one hand. All fingers on the affected hand were examined, to determine if sweat and sensation were present or absent. Because of the physiology of the situation, only twelve observations were recorded on the ring finger. The response of the ring finger is affected by two nerves (the median and ulnar), while the other fingers are innervated only by one. In order for the ring finger to be desensitized, both nerves had to be impaired. The observations are encoded as "Agree with hypothesis," "Disagree with hypothesis." Table 1 gives these data.

---

[1]Department of Biostatistics, UNC Chapel Hill

[2]Hand Rehabilitation Center, UNC Chapel Hill

TABLE 1

DATA FROM SWEAT AND SENSATION EXPERIMENT

	Thumb	Index	Long	Ring	Fifth	Total
Agree	25	30	27	10	29	121
Disagree	10	5	8	2	6	31
Proportion Disagree	.286	.143	.229	.167	.171	.203

Clearly the results on one finger cannot be assumed independent of other fingers. Thus the data cannot be treated as a contingency table with a fixed marginal. The problem of how to test for homogeneity of proportions remains. It may also be useful to test if the proportion disagreeing is 0 (or perhaps some small value such as .01). The correspondence between confidence intervals and hypothesis testing provides a solution to the problem.

Goodman [1,2] has studied the case of independent observations on multinomial variables and gives an asymptotic solution to the problem. He also indicates that an approach similar to the one suggested here may give shorter intervals when the number of comparisons is small. This paper gives a procedure which may be used quite generally and in particular is applicable to dependent observations. The approach of Goodman [2] might have been used on this example by converting the five binomial variates into $2^5=32$ states. Unfortunately, this procedure would require that an observation be available from each finger on each person (i.e. no missing values) which simply was not possible in this situation.

Miller [4] has given a survey of simultaneous inference. In this book he discusses the Bonferroni inequalities with special reference to the problem of testing hypotheses. A very attractive feature of these inequalities is that they may be used for dependent, as well as independent, variables. In the context of our problem, let E_i be the event that P_i lies in a random interval, say (ℓ_i, u_i).

The event that all P_i lie in their respective intervals will be denoted $A = \bigcap_i E_i$. The Bonferroni inequality states

$$\Pr\{A\} \geq 1 - \sum \Pr\{\sim E_i\}.$$

If α_i is the probability that the i^{th} interval does not cover P_i than we have

$$\Pr\{A\} \geq 1 - \sum \alpha_i.$$

In words, the probability that all intervals cover P_i is greater than or equal to one minus the sum of the probabilities that each interval does not cover its P_i. A simple choice of α_i is $\alpha_i = \alpha/k$ where k is the number of groups. When k is not large, these inequalities are fairly good. Lachenbruch [3] has given tables simultaneous confidence intervals for the Binomial and Poisson distributions based on the Bonferroni inqualities. Using these tables we find the .90 and .95 confidence intervals which are given in table 2.

TABLE 2

CONFIDENCE INTERVALS FOR DATA OF TABLE 1

	Thumb	Index	Long	Ring	Fifth
.90 interval	(.127,.494)	(.038,.333)	(.088,.432)	(.013,.537)	(.054,.367)
.95 interval	(.115,.515)	(.032,.354)	(.079,.454)	(.009,.573)	(.046,.389)

To test H_o: $P_1 = P_2 = P_3 = P_4 = P_5$

against H_1: $P_i \neq P_j$ for at least one pair $i \neq j$

we note that if a pair of intervals overlap the corresponding test will not reject the hypothesis that the pair of P_i's are the same. In our problem, all intervals overlap so the hypothesis of homogeneity of the P_i's cannot be rejected at either the .10 or .05 significance level.

Next, suppose we wish to test the hypothesis that $\underset{\sim}{P} = (P_1 \ldots P_5)'$ is equal to a vector of small values, say H_o: $\underset{\sim}{P} = (.01, .01, .01, .01, .01)'$. We note that if any confidence interval fails to cover the value specified by H_o (in this case .01) we reject H_o. Thus, at the α=.05 level, we use the .95 confidence intervals to find that the null hypothesis is rejected. The individual intervals indicate that only for the ring finger is there insufficient evidence to reject.

As k, the number of groups, becomes large, the Bonferroni intervals become wide and loss of power results. Therefore, this approach does not appear to be a good one for large k. The intervals in (3) are given for $k \le 10$. The tables referred to also make it possible to perform the same analysis on Poisson variables.

A natural extension of this procedure is to examine linear contrasts of the P_i. Lemma 1, p.44 of Miller [4] enables us to show that if $|\hat{P}_i - P_i| \le d$ for all i then

$$\left|\sum c_i(\hat{P}_i - P_i)\right| \le d \sum |c_i|$$

Recall that E_i was defined as the event $\ell_i \le P_i \le u_i$. If E_i is true then $|\hat{P}_i - P_i| \le u_i - \ell_i$. For our purposes, we may choose $d = \max_i (u_i - \ell_i)$. Then we have

$$\Pr\left\{\left|\sum c_i(\hat{P}_i - P_i)\right| \le d \sum |c_i|\right\} \ge 1-\alpha$$

If $H_o: \sum c_i P_i = 0$ holds, then we can construct a test of H_o of size α: Accept H_o if and only if

$$\left|\sum c_i P_i\right| \le d \sum |c_i| \ .$$

Confidence intervals for linear contrasts can be obtained in the usual manner.

There are several competitors of the intervals presented here for the binomial case. Not so much is known about the Poisson case. The method given here may be used for any distribution for which simultaneous intervals are available.

For purposes of comparison, we use three methods suggested by Goodman. First, we have k multinomial populations with J classes each. Let P_{ij}, $i=1,\ldots,k$, $j=1,\ldots,J$ be the corresponding cell probabilities and define a contrast $\Sigma c_{ij} P_{ij}$ where the c_{ij} are known constants subject to $\sum_i c_{ij} = 0$ for $j=1,\ldots,J$. Because of the restrictions of the problem there are $(k-1)(J-1)$ linearly independent contrasts. For the binomial case J=2 and there are exactly k-1 independent contrasts. Goodman (1) showed that for any contrast, say $A = \Sigma c_{ij} P_{ij}$, a $100(1-\alpha)\%$ simultaneous confidence interval was given by

$$\hat{A} - S_{\hat{A}} L \leq A \leq \hat{A} + S_A L$$

where

$$\hat{A} = \sum_{i,j} c_{ij} \hat{P}_{ij} = \underset{\sim}{C}' \underset{\sim}{\hat{P}}$$

$$S_{\hat{A}}^2 = \sum_{i=1}^{k} \frac{1}{n_i} \left[\sum_{j=1}^{J} c_{ij}^2 P_{ij} - \left(\sum_{j=1}^{J} c_{ij} P_{ij} \right)^2 \right].$$

$$= \sum_i \frac{1}{n_i} [\underset{\sim}{C}_i' V_i \underset{\sim}{C}_i]$$

and $L^2 = 100(1-\alpha)$ percentile point of the $\chi^2_{(k-1)(J-1)}$ distribution. $\underset{\sim}{C}' = (\underset{\sim}{C}_1', \underset{\sim}{C}_2', \ldots, \underset{\sim}{C}_k')$ and Σ_i is the covariance matrix in the i<u>th</u> population. Note that this method does not require equal sample sizes from the k multinomial probabilities. It does not permit estimation of individual probabilities as they can't be expressed as contrasts. This method will be denoted as the Goodman χ^2 (GC) method.

In a later paper (2), Goodman considered problems of a single population and gave intervals for all linear combinations (not contrasts). If there are J categories, then a 100(1-α) percent confidence interval for the linear combination $\underset{\sim}{\ell}'$ is

$$\underset{\sim}{\ell}'\underset{\sim}{P} \pm (\chi^2_{J-1})^{\frac{1}{2}}(\ell'\hat{V}_P\ell/n)^{\frac{1}{2}}.$$

If we specialize the case of several binomial populations to the case of equal sample size, we may fit it into this model. This method will be called the Extended Goodman χ^2 (EGC) method.

Note that we may relate a set of k binomial distributions (independent or dependent) with equal n to a single multinomial distribution with 2^k states. The k binomial probabilities may be obtained from k linear combinations (not contrasts). If we denote the transformation matrix by D, we see that D is an $k \times 2^k$ matrix consisting of 0's and 1's. For example, let I=3, the possible responses are (0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0) and (1,1,1). If the $j^{\underline{th}}$ one of these has probability π_j and we wish to find P_i, the probability of success in the $i^{\underline{th}}$ component of these triplets, we have

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

the covariance matrix of the $\underset{\sim}{P}$'s is

$$V_{\underset{\sim}{P}} = DV_\pi D'$$

where V_π is the usual multinomial covariance matrix.

We may write a contrast between the $\underset{\sim}{P}$'s as $\underset{\sim}{C}'\underset{\sim}{P} = \underset{\sim}{C}'D\underset{\sim}{\Pi}$. The EGC method yields the following interval for the contrast $\underset{\sim}{C}'\underset{\sim}{P}$.

$$\underset{\sim}{C}'\underset{\sim}{P} \in \underset{\sim}{C}'\underset{\sim}{P}^* \pm (\chi^2_{2^k-1})^{1/2}(\underset{\sim}{C}'V^*_{\underset{\sim}{P}}\underset{\sim}{C})^{1/2}$$

where $\underset{\sim}{P}^*$ is the maximum likelihood estimator of $\underset{\sim}{P}$, $\underset{\sim}{C}'V_{\underset{\sim}{P}}\underset{\sim}{C}$ is the variance of $\underset{\sim}{C}'\underset{\sim}{P}^*$ and $\chi^2_{2^k-1}$ is the 100(1-α) per cent point of the χ^2 distribution with 2^k-1 degrees of freedom. These intervals are unnecessarily wide if only a limited number of comparisons, say m, are to be performed. In this case Goodman suggests using g_m, the Bonferroni critical value for the normal distribution instead of the χ^2 value. We shall assume the P_i's are independent. Then, the variance of $\underset{\sim}{C}'\underset{\sim}{P}^*$ is

$$\sum C_i^2 P_i(1-P_i)/n \leq \sum C_i^2 \quad \max_i \frac{P_i(1-P_i)}{n}$$

and a conservative set of intervals is

$$\underset{\sim}{C}'\underset{\sim}{P} \in \underset{\sim}{C}'\underset{\sim}{P}^* \pm (\chi^2_{2^k-1})^{1/2}(\sum C_i^2) \quad \max\left(\frac{P_i(1-P_i)}{n}\right)^{1/2}$$

or

$$\underset{\sim}{C}'\underset{\sim}{P} \in \underset{\sim}{C}'\underset{\sim}{P}^* \pm g_m \, (\sum C_i^2)^{1/2} \max\left(\frac{P_i(1-P_i)}{n}\right)^{1/2}.$$

The intervals proposed in this paper are (for large n)

$$\underset{\sim}{C}'\underset{\sim}{P} \in \underset{\sim}{C}'\underset{\sim}{P}^* \pm g_k \quad \sum |C_i| \quad \max\left(\frac{P_i(1-P_i)}{n}\right)^{1/2}.$$

In general, we know that

$$(\sum c_i^2)^{\frac{1}{2}} \leq \sum |c_i| \qquad \text{since}$$

$$(\sum |c_i|)^2 = \sum c_i^2 + 2 \sum_{i>j} |c_i| \, |c_j| > \sum c_i^2.$$

So we may anticipate that the Goodman intervals will be shorter than the ones proposed here. Table 3 compares the methods at the α=.05 level for independent variables, assuming that the comparison of interest is the longest interval. Thus, we give the coefficient of $\max(\frac{P_i(1-P_i)}{n})^{\frac{1}{2}}$. In a sense, this is unfair to the Goodman intervals which are not designed to minimize the maximum interval length. However, it is difficult to envisage what a typical standard deviation would be without assuming a prior distribution on the P_i's. Helmert's contrasts which are included in the table are the familiar set of orthogonal contrasts which are given by

$$\underset{\sim}{c}_1' = (1, \ -1, \ \ 0, \ \ 0, \ \ 0)$$

$$\underset{\sim}{c}_2' = (1, \ \ 1, \ -2, \ \ 0, \ \ 0)$$

$$\underset{\sim}{c}_3' = (1, \ \ 1, \ \ 1, \ -3, \ \ 0)$$

$$\underset{\sim}{c}_4' = (1, \ \ 1, \ \ 1, \ \ 1, \ -4)$$

for k=5. They have been normalized so that $\sum c_i^2=1$. The values given in parentheses are the relative efficiencies of the LP method to the method under consideration.

The GC intervals are not available for intervals for particular P_i, and so cannot be compared to the LP method for this combination. For small k, it appears that the GC method will usually give shorter intervals than the LP method. In only one case (all pairwise contrasts for k=10) was the LP better than the GC. The EGC method allows one to estimate intervals for various probabilities, but the price is rather expensive. The intervals are quite wide, and always much larger than the LP method. Table 4 gives approximate values of χ^2 to be used with this method. The approximation used is $\chi^2_\alpha = \nu(1 - 2/9\nu + z_\alpha \sqrt{2/9\nu})^3$ where z_α is $100\alpha^{\underline{th}}$ percentile of the normal distribution.

The GB intervals are shorter than the GC or LP intervals as might be expected. The GB intervals are designed for a pre-planned set of contrasts while the others are valid for all contrasts. If a very large number of contrasts are planned, the GB intervals might be longer than the LP intervals, but such a situation seems unlikely to arise in practice. The LP vs EGC comparison indicates a substantial difference in favor of the LP method, particularly for large k. The LP-GC comparison is less clear cut. We would recommend the LP when intervals for means are desired or when I is large and pairwise contrasts are of interest; otherwise, the GC method is better. It should be noted that the EGC method is not applicable for unequal sample sizes.

The efficiency values indicate that the performance of the LP method relative to the other ones depends on the contrasts that are to be used. If the contrasts were known in advance, then the assessment would be easy, but in this situation, the GB method will usually be best. This seems to leave us in the position of deciding not only the contrasts to be examined, but also the method of comparison, after the data has been collected. We do not know of any studies on this topic.

Finally, the LP method can be easily extended to the Poisson case. The Goodman methods are not applicable in this situation.

Table 3

Comparison of Maximum Interval Length for
Various Linear Combinations (α=.05)
(efficiency)

k	Comparison	Number	Goodman χ^2	Extended Goodman χ^2	Goodman Bonferroni	Lachenbruch-Perry
5	All means	5	——— (———)	6.71 (6.76)	2.58 (1.00)	2.58
5	All Pairwise Contrasts	10	4.36 (.71)	9.49 (3.39)	3.97 (.59)	5.16
5	Helmert's Contrasts	4	3.08 (.44)	6.71 (2.10)	2.50 (.29)	$5.16\sqrt{i-1/i} \leq 4.62$
10	All means	10	——— (———)	33.15 (139.2)	2.81 (1.00)	2.81
10	All Pairwise Contrasts	45	5.81 (1.06)	46.87 (69.6)	4.61 (.67)	5.62
10	Helmert's Contrasts	9	4.11 (.59)	33.15 (38.6)	2.77 (.27)	$5.62\sqrt{i-1/i} \leq 5.34$

Table 4

Approximate Values of χ^2 for 2^k-1 df

df	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.99}$
31	41.4	45.0	52.2
63	77.7	82.5	92.0
127	147.8	154.3	167.0
255	284.3	293.3	310.5
511	552.4	564.7	588.3
1023	1081.4	1098.5	1131.1

References

1. Goodman, L. A. (1964) "Simultaneous Confidence Limits for Contrasts Among Multinomial Populations" Ann. Math. Statist. v.35, p.716

2. Goodman, L. A. (1965) "On Simultaneous Confidence Intervals for Multinomial Proportions" Technometries v.7, p.247

3. Lachenbruch, P. A. (1968) "Simultaneous Confidence Limits for the Binomial and Poisson Distributions "Institute of Statistics Mimeo Series No. 596

4. Miller, R. G. (1966) Simultaneous Statistical Inference New York: McGraw-Hill