Relationship Between Smoking and Calculus Deposition

CHARLES J. KOWALSKI

University of Michigan School of Dentistry, Ann Arbor, Michigan 48104, USA

The chi-square test for association in contingency tables is often used to analyze data gathered during the course of dental/oral experimentation. A simple procedure that makes such analyses more useful is described, and the method is applied in a study of the relationship between smoking and calculus deposition.

An extensive study of the relationship between smoking and calculus deposition in nearly 6,000 men was undertaken by Pindborg,[1] in 1949. Chilton,[2] in 1967, used these data to illustrate the familiar test for association in a contingency table but noted that the analysis was limited because, although it established the existence of a significant association between smoking and calculus deposition, it did not answer the important question of which particular groups (or combinations of groups) differed significantly. He concluded that "what is needed here are analogues of the tests* suggested by Tukey[3] and by Scheffé,"[4] but apparently he failed to realize that such analogues are, in fact, available and may easily be used to extract additional valuable information from contingency table analyses. Two procedures that accomplish this extended analysis and illustrate the power of the method in the analysis of Pindborg's[1] data are outlined in this study.

Materials and Methods

Two asymptotic (large sample) methods for making multiple comparisons among the probabilities in an $R \times C$ contingency table, reported by Gold,[6,7] are apparently

relatively unknown and inaccessible to dental/oral researchers. These are presented by giving the data and carrying the analysis as far as Chilton[2] did to motivate the developments to follow.

The table summarizes the data from Pindborg's[1] study. Chilton[2] used the chi-square test to confirm the null hypothesis that smoking is not related to calculus formation. The observed value of chi-square (with four degrees of freedom) is $\chi_4^2 = 79.71$, and this is significant at the 0.01 level $(P < 0.01)$; thus, an association between smoking and calculus deposition is then accepted. But this is all the chi-square test can tell us; it cannot be used to discover where the significance lies. Chilton[2] claimed that "it can be seen that nonsmoking runs more freely with no calculus, and heavy smoking (>10 grains) runs more freely with calculus formation. In addition, subgingival calculus seems to be affected less by no smoking or heavy smoking than the two other categories of no calculus and supragingival calculus." But these conclusions are based on a visual inspection of the data. What is needed are methods for formally testing these hypotheses.

The first method described in this study can be used to test any and all hypotheses of this type, including those suggested by the data (so this is the analogue of the Tukey[3] and Scheffé[4] procedures), whereas the second method is appropriate only for hypotheses stated before the collection and initial analysis of the data (this is analogous to Dunn's[8] modification of the Tukey[3] and Scheffé[4] procedures).

The description of the methods is somewhat general, to facilitate application to problems other than the one considered here. However, the abstract concepts will be identified with the concrete quantities in the table for clarity.

* These tests are used after a significant analysis of variance to pinpoint which of the means (or combinations of means) are significantly different. A discussion may be found in Chilton's book.[2] The term multiple comparisons is frequently used to describe such procedures. A complete account is given by Miller.[5]

TABLE
OBSERVED DISTRIBUTION OF 5,690 MEN BY SMOKING HABITS AND
CALCULUS DEPOSITION

Subjects	No Calculus	Supragingival Calculus	Subgingival Calculus	Totals
Nonsmokers	284 (0.5000)*	236 (0.4155)	48 (0.0845)	568
Smokers < 10 grains per day	606 (0.3370)	983 (0.5467)	209 (0.1162)	1,798
Smokers > 10 grains per day	1,028 (0.3093)	1,871 (0.5629)	425 (0.1279)	3,324
Totals	1,918	3,090	682	5,690

* Proportions are given in parentheses.

METHOD 1: SIMULTANEOUS CONFIDENCE INTERVALS FOR ALL LINEAR FUNCTIONS OF CELL PROBABILITIES.—Let $P_1, \ldots, P_R$ be R populations (smoking habits in the example), each with C categories (calculus deposition), and let p_{ij} $(i = 1, \ldots, R: j = 1, \ldots, C)$ be the (unknown) probability of the jth category in P_i. For example, P_{12} denotes the probability that a nonsmoker has supragingival calculus. Let n_i be the number of observations from P_i (the row totals in the table), and denote by x_{ij} the number of observations in the jth category in the sample from P_i (the numbers in the cells of the table). The method described here gives simultaneous confidence intervals for all linear functions:

$$\theta = \sum_{i=1}^{R} \sum_{j=1}^{C} a_{ij} p_{ij}, \quad (1)$$

where a_{ij} may be any real numbers (usually a_{ij} are contrasts, ie, they are chosen so that their sum is zero, eg, $p_{11} - p_{12}$ or $p_{11} - \frac{1}{2}(p_{12} + p_{13})$, but this is not a necessary limitation). Equation 1 is estimated by

$$\hat{\theta} = \sum_{i=1}^{R} \sum_{j=1}^{C} a_{ij} \hat{p}_{ij}, \quad (2)$$

where

$$\hat{p}_{ij} = x_{ij}/n_i \quad (3)$$

are the proportions in the cells of the table. The variance of $\hat{\theta}$ is

$$\sigma^2(\hat{\theta}) = \sum_{i=1}^{R} \frac{1}{n_i} \left[\sum_{j=1}^{C} a_{ij}^2 p_{ij} - \left(\sum_{j=1}^{C} a_{ij} p_{ij} \right)^2 \right], \quad (4)$$

and this is estimated by

$$S^2(\hat{\theta}) = \sum_{i=1}^{R} \frac{1}{n_i} \left[\sum_{j=1}^{C} a_{ij}^2 \hat{p}_{ij} - \left(\sum_{j=1}^{C} a_{ij} \hat{p}_{ij} \right)^2 \right]. \quad (5)$$

In this notation, Gold[6] proved that, if the n_i are large, the probability is at least 1-α that the values of all functions θ simultaneously satisfy

$$\hat{\theta} - KS(\hat{\theta}) \leqq \theta \leqq \hat{\theta} + KS(\hat{\theta}), \quad (6)$$

where K is the positive square root of the upper $(1\text{-}\alpha)^{\text{th}}$ percentage point of the chi-square distribution with $R(C\text{-}1)$ degrees of freedom. Thus, if we take $\alpha = 0.05$, the upper 95% point of the χ_6^2 distribution is 12.6 and so $K = \sqrt{12.6} = 3.548$.

METHOD 2: SIMULTANEOUS CONFIDENCE INTERVALS FOR PREDETERMINED LINEAR COMBINATIONS OF CELL PROBABILITIES.— Method 1 applies here, but now we must prespecify a number, G, of linear combinations of the form of equation 1 for which we desire simultaneous confidence intervals. Symbolically, we want confidence intervals for G linear combinations:

$$\theta_g = \sum_{i=1}^{R} \sum_{j=1}^{C} a_{ij}^{(g)} p_{ij} \quad (g = 1, \ldots, G), \quad (7)$$

where $a_{ij}^{(g)}$ are G sets of prespecified constants. Equations 2 through 5 continue to hold, with the obvious modification in notation: all that changes is the value of the constant in equation 6. In this instance, provided that n_i are large, the probability is

at least $1-\alpha$ that the values of G functions θ_g simultaneously satisfy

$$\hat{\theta}_g - Z_G S(\hat{\theta}_g) \leqq \theta_g \leqq \hat{\theta}_g + Z_G S(\hat{\theta}_g), \quad (8)$$

where Z_G is the $(1-\alpha/2G)^{\text{th}}$ upper percentage points of the standard normal distribution. In particular, it should be noted that Z_G depends on G, the number of confidence intervals required. If, eg, $G = 12$ and 95% confidence intervals are desired ($\alpha = 0.05$), the appropriate constant is $Z_{12} = 2.865$. If, however, $G = 6$ intervals were sufficient, $Z_6 = 2.368$, and the intervals would be narrower, but only half the number of intervals could be studied. Thus, when choosing G, the investigator should attempt to include all the intervals in which he is interested but not those that are unlikely to be of practical importance.

Results

Application of these methods to the data of the table may now be considered.

METHOD 1.—With method 1, it is possible to obtain a confidence interval for any function of the form equation 1, regardless of whether it is suggested by the data. If we wanted to study the difference, among nonsmokers and moderate smokers, of the probabilities of having no calculus, we would consider $\theta = p_{11} - p_{12}$ (this corresponds to the choice of $a_{11} = 1$, $a_{12} = -1$, all other $a_{ij} = 0$ in equation 1 and is a contrast, since $a_{11} + a_{12} = 0$). This is estimated by $\hat{\theta} = 0.5000 - 0.3370 = 0.1630$. The estimated variance (5) is $S^2(\hat{\theta}) = 0.0005644$ and, hence, $S(\hat{\theta}) = 0.02376$. For a 95% confidence interval, we use $K = 3.548$. Then, by equation 6, a 95% confidence interval for $\theta = p_{11} - p_{12}$ is (0.0787, 0.2473). Since this interval does not include zero, we may reject (at the 5% level of significance) the equality of p_{11} and p_{12} and, hence, conclude that nonsmokers have a significantly higher probability of having no calculus than of having supragingival calculus. This clarifies one aspect of the statement that "nonsmoking runs more freely with no calculus," and the corresponding arithmetic penance is perhaps not too severe in view of the sharpness of the result.

Any other hypothesis of this type, however evolved, may be tested by use of this method. To test that the probability of supragingival calculus among nonsmokers is the same as the average of this probability in the two smoking groups, one considers the contrast $\theta = p_{12} - \frac{1}{2}(p_{22} + p_{32})$. The 95% confidence interval is $(-0.2171, -0.0615)$, and the conclusion is that nonsmokers have a smaller probability of supragingival calculus than does the average smoker.

METHOD 2.—It often happens that the investigator is not concerned with all possible linear combinations of the type equation 1 and is able to prespecify functions to which he is willing to confine his analysis. The smaller the number G, the smaller the resultant confidence intervals (ie, the more sensitive the tests), but the analysis must be restricted (to maintain the specified error rate α) to only these G comparisons.

For example, suppose the investigator settles on $G = 12$ comparisons, including $p_{11} - p_{12}$. In this instance, $Z_{12} = 2.865$ and the 95% confidence interval is (0.0949, 0.2311). This is a tighter interval than that obtained by use of method 1, but the investigator must look only at the 12 prespecified comparisons. This is the crux of the choice between which method should be used in practice. The 95% confidence interval for the contrast $\theta = p_{12} - \frac{1}{2}(p_{22} + p_{32})$, if this were included among the 12 prespecified comparisons, would be $(-0.2021, -0.0765)$, again a smaller interval than that obtained by method 1.

For the sake of completeness, some nonsignificant comparisons, ie, contrasts the confidence intervals of which include the value zero, are given. Even these results often contain valuable information: eg, the 95% confidence intervals (either method of construction) for the three contrasts $p_{21} - p_{31}$, $p_{22} - p_{32}$, and $p_{23} - p_{33}$ all contain zero. The conclusion is that the two groups of smokers are homogeneous with respect to calculus deposition. Otherwise stated, it is not so much the amount smoked that leads to differential calculus formation, but rather the fact of smoking: The amount smoked is unlikely to be a good indicator of calculus formation.

Discussion

The two methods described do not exhaust the number of solutions to the problem posed that appear in the statistical lit-

erature. Other approaches were taken by Quesenberry and Hurst[9] and Goodman.[10,11] Summaries of these methods may be found in Miller.[5] Gold's[6] method was used mainly because of its simplicity. Although other methods may, in some instances, produce even narrower confidence intervals, the methods described here are particularly easy to apply and at least illustrate the fact that analogues of the Tukey[3] and Scheffé[4] procedures that may provide considerable additional insight into the structure of the data are available.

It should be noted that these procedures are strictly valid only as $n_i \rightarrow \infty$ and so should be used with caution if n_i are small. Of course, if n_i are small, the chi-square test itself is only approximate,[12] so that this point need not cause undue concern in practice. We are on solid ground in the particular example we have been considering, but the procedures may be used on a number of other problems as well.

Conclusions

According to the analyses described, the chances of having no calculus, supragingival calculus, and subgingival calculus do not differ in the two groups of smokers studied. It was not so much the amount smoked, but rather the fact of smoking that influenced calculus formation. The probability of being calculus-free was greater for nonsmokers than for either group of smokers, whereas the probabilities of supragingival and subgingival calculus among nonsmokers were less than for the average smoker. Subgingival calculus was affected less by no smoking or heavy smoking than the two other categories, no calculus and supragingival calculus. The methods described allow formal tests of these hypotheses, ie, the conclusions reached are backed up by rigorous probability statements.

References

1. PINDBORG, J.J.: Tobacco and Gingivitis: II. Correlation Between Consumption of Tobacco, Ulceromembranous Gingivitis, and Calculus, *J Dent Res* **28**:461, 1949.
2. CHILTON, N.W.: *Design and Analysis in Dental and Oral Research*, Philadelphia: J. B. Lippincott Co., 1967, pp 218-225.
3. TUKEY, J.W.: *The Problem of Multiple Comparisons*, Princeton, NJ: Princeton University Press, 1953.
4. SCHEFFÉ, H.: A Method of Judging All Contrasts in the Analysis of Variance, *Biometrika* **40**:87, 1953.
5. MILLER, R.G.: *Simultaneous Statistical Inference*, New York: McGraw-Hill, 1966.
6. GOLD, R.Z.: Tests Auxiliary to χ^2 Tests in a Markov Chain, *Ann Math Stat* **34**:56, 1963.
7. BERGER, A., and GOLD, R.Z.: On Comparing Survival Times, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, 1961, p 67.
8. DUNN, O.J.: Multiple Comparisons Among Means, *J Amer Stat Assn* **56**:52, 1961.
9. QUESENBERRY, C.P., and HURST, D.C.: Large Sample Simultaneous Confidence Intervals for Multinomial Proportions, *Technometrics*, **6**:191, 1964.
10. GOODMAN, L.A.: Simultaneous Confidence Intervals for Contrasts Among Multinomial Proportions, *Ann Math Stat* **35**:716, 1964.
11. GOODMAN, L.A.: On Simultaneous Confidence Intervals for Multinomial Proportions, *Technometrics*, **7**:247, 1965.
12. CHILTON, N.W.: *Design and Analysis in Dental and Oral Research*, Philadelphia: J. B. Lippincott Co., 1967, pp 225-226.