# ANACONDA: ANALYSIS OF CONCORDANCE OF g GROUPS OF JUDGES

James Beckett III* and William R. Schucany**
Southern Methodist University

## INTRODUCTION

The general problem considered is the analysis of the agreement within and between several groups of judges. For example, we may have Black, White, and Mexican/American consumers assigning the ranks 1 through k to k products as to their preference. The $\mathcal{L}$ statistic is used to examine the agreement between any two groups of judges. One group may have m judges with object rank totals $S_j(j = 1,2,\cdots,k)$, while the second group may have n judges with object rank totals $T_j(j = 1,2,\cdots,k)$. Schucany (1971) defined the statistic

$$\mathcal{L} = \sum_{j=1}^{k} S_j T_j = \underset{\sim}{S}'\underset{\sim}{T}$$ where $\underset{\sim}{S}$ and $\underset{\sim}{T}$ are the vectors of the column sums of ranks for each of the two groups. The elements of $\underset{\sim}{S}$ and $\underset{\sim}{T}$ are

$$S_j = \sum_{i=1}^{m} R_{ij},\ j = 1,2,\cdots,k, \text{ and } T_j = \sum_{i=1}^{n} R'_{ij},$$

$j = 1,2,\cdots,k$, where $R_{ij}(R'_{ij})$ represents the rank given the $j^{th}$ object by the $i^{th}$ judge in group one (two). With rank data on two groups there are two questions of practical importance. Do the groups agree, and, if so, what is the combined consensus ranking of the k products or objects? Before taking up the partitioning of a measure of the overall concordance of g groups of rankings, we shall discuss a technique for comparing objects within a set which are not all indistinguishable.

## MULTIPLE COMPARISON THEORY

After a significant $\mathcal{L}$, which indicates agreement between groups as well as agreement within each group, we may desire to compare some or all of the objects. For example two groups of tasters may be ranking soft drinks for taste. Our primary concern may be the ordering of only two of the products, e.g., Coke vs. Pepsi. How do we perform even a preplanned comparison? Some appropriate yardstick is needed, i.e., some measure of the difference or similarity of two objects.

Although $\mathcal{L}$ is made up of the sum of the rank total products, using these rank total products individually (as a measure of the degree of preference for that object) proves to be misleading and inappropriate. So instead of multiplying the object rank totals, we add them. More generally we take the linear combination $aS_i + bT_i$, a weighted rank total enabling us to weight judges or groups. For example, with Page's L, one group of judges versus an expert, a and b may be chosen so that the experts' opinion is weighted equal to one other judge or equal to all the other judges. With $\mathcal{L}$ the most common weighting schemes would be $a = b = 1/2$ (each judge equal voice) or $a = \frac{n}{m+n}$ and $b = \frac{m}{m+n}$ (each group equal voice).

Given a and b, we proceed in much the same way as Nemenyi approaches the one-group (Friedman) multiple comparison problem [see Miller (1966)]. Using the *known* standard deviation of $aS_i + bT_i$, i.e. (error degrees of freedom) $\nu = \infty$, we apply a Duncan multiple range test to the resultant adjusted "treatment means".

## EXAMPLE

Table 1 is a tabulation of the ranking of four wines by six Frenchmen and nine Americans. A (significant) value of $\mathcal{L}$ * (standard normal) equal to 4.90 is obtained for this contrived [Schucany and Frawley (1973)] example. The proper conclusion is that the Frenchmen and Americans are in agreement as to the preference rankings of wines A, B, C, and D. The (weighted object rank total) vector $\underset{\sim}{C}^*$ is defined as

$$\underset{\sim}{C}^* = \frac{a\underset{\sim}{S} + b\underset{\sim}{T}}{\sqrt{\frac{k(k+1)}{12}(a^2m + b^2n)}},$$

where a and b are the aforementioned nonnegative weighting constants with $a + b = 1$. The constant denominator in $\underset{\sim}{C}^*$ is the known standard deviation of $aS_i + bT_i$.

Using $a = b = 1/2$, i.e., all judges have equal influence, we obtain $\underset{\sim}{C}^{*\prime} = (5.2\ \ 5.6\ \ 8.4\ \ 10.8)$. We compare wines A through D by comparing differences of the $C_j^*$ to the percentage points of the Duncan Multiple Range Test, using $\nu = \infty$ and $\alpha = .05$. Displaying the results we have

| A | B | C | D |
|---|---|---|---|
| 5.2 | 5.6 | 8.4 | 10.8 |

(A and B underlined together; C and D underlined together.)

Our conclusion is that the two groups of judges are in agreement on the ordering of the wines and they agree that wines A and B are better (worse) than wines C and D.

TABLE 1

Independent Rankings of Wine Vintages by Wine Connoisseurs

| | 6 Rankings | WINES A | B | C | D |
|---|---|---|---|---|---|
| FRENCH JUDGES | 1 | 2 | 1 | 3 | 4 |
| | 2 | 1 | 3 | 4 | 2 |
| | 3 | 1 | 3 | 2 | 4 |
| | 4 | 1 | 4 | 2 | 3 |
| | 5 | 3 | 1 | 2 | 4 |
| | 6 | 1 | 2 | 4 | 3 |
| | Totals | 9 | 14 | 17 | 20 |

| | WINES | | | | |
|---|---|---|---|---|---|
| | 9 Rankings | A | B | C | D |
| AMERICAN JUDGES | 1 | 1 | 2 | 4 | 3 |
| | 2 | 3 | 1 | 2 | 4 |
| | 3 | 1 | 2 | 3 | 4 |
| | 4 | 1 | 2 | 3 | 4 |
| | 5 | 3 | 2 | 1 | 4 |
| | 6 | 2 | 1 | 4 | 3 |
| | 7 | 1 | 2 | 3 | 4 |
| | 8 | 2 | 1 | 3 | 4 |
| | 9 | 3 | 1 | 2 | 4 |
| | Totals | 17 | 14 | 25 | 34 |

## ANACONDA

The concept of ANACONDA (the shortened form of Analysis of Concordance) is similar to the concept of ANOVA; indeed ANACONDA bears the same relationship to the $\mathcal{L}$ test as ANOVA to the t-test. A one-way classification of judges into three or more groups or subpopulations is a frequent situation. As evidenced by the SP (Sum of Products) column of Table 2, the total agreement is partitioned into the various sources of agreement. Li and Schucany (1975) exposed the false conclusions which might occur by looking only at the combined (pooled groups) Total chi-square. Nevertheless many practitioners use only the individual group and pooled group chi-squares leaving out the $\mathcal{L}$ statistics which truly measure the agreement between groups. These $\mathcal{L}$ statistics should be examined first in analyzing a One-Way ANACONDA Table such as Table 2. We must examine the individual group chi-squares only if at least one of the $\mathcal{L}$ statistics is not significant. Theoretical results [Beckett (1975)] have established the zero correlation and asymptotic independence between any two $\mathcal{L}$'s and between one of the $\mathcal{L}$ statistics and a related Friedman $\chi^2$.

Table 3 illustrates a two-way (2 x 2) ANACONDA Table. The rightmost column is added to aid in interpretation. Analysis begins with the examination of the significance of Factor 1 and Factor 2.

Their significance along with the significance of $\mathcal{L}_I$ allows the conclusion that all subpopulations are in agreement. If at least one of the main effect $\mathcal{L}$'s is not significant, the respective Friedman $\chi^2$'s are investigated.

## SUMMARY

The determination of a consensus ranking of the products based on the rankings of each judge can be accomplished through the use of a "Duncan"-type multiple comparison procedure. Finally the extension from two groups to g groups of judges is simply presented as an ANACONDA (Analysis of Concordance) Table.

## FOOTNOTES

* Now at Bowling Green State University

** Research partially supported by ONR contract N00014-75-C-0439

## REFERENCES

[1] Beckett, J. (1975). "Some Properties and Applications of a Statistic for Analyszing Concordance of Rankings of Groups of Judges." Unpublished dissertation, Southern Methodist University.

[2] Li, L. and Schucany, W.R. (1975). "Some Properties of a Test for Concordance of Two Groups of Rankings," Biometrika, 62, 417-423.

[3] Miller, R.G. (1966). Simultaneous Statistical Inference. New York: McGraw-Hill Book Company.

[4] Page, E.B. (1963). "Ordered hypothesis for multiple treatments: a significance test for linear ranks," JASA, 58, 216-230.

[5] Schucany, W.R. (1971). "A rank test for two group concordance", Annals of Mathematical Statistics (Abstract), 42, 1146.

[6] Schucany, W.R. and Frawley, W.H. (1973). "A rank test for two group concordance," Psychometrika, 38, 249-258.

TABLE 2

One-Way ANACONDA TABLE

| Source | SP | Statistic |
|---|---|---|
| Black | $\underline{B}'\underline{B}$ | $\chi^2_B$ |
| White | $\underline{W}'\underline{W}$ | $\chi^2_W$ |
| Mex./Am. | $\underline{M}'\underline{M}$ | $\chi^2_M$ |
| Black vs. White | $\underline{B}'\underline{W}$ | $\mathcal{L}_{BW}$ |
| Black vs. Mex./Am. | $\underline{B}'\underline{M}$ | $\mathcal{L}_{BM}$ |
| White vs. Mex./Am. | $\underline{W}'\underline{M}$ | $\mathcal{L}_{WM}$ |
| All Groups | $\underline{B}'\underline{W} + \underline{B}'\underline{M} + \underline{W}'\underline{M}$ | $\mathcal{L}^*_{BW} + \mathcal{L}^*_{BM} + \mathcal{L}^*_{WM}$ |
| TOTAL | $(\underline{B}+\underline{W}+\underline{M})'(\underline{B}+\underline{W}+\underline{M})$ | $\chi^2_C$ |

TABLE 3

Two-Way ANACONDA (2 x 2)

| Source | SP | Statistic | Interpretation |
|---|---|---|---|
| Black Men | $\underline{S}'\underline{S}$ | $\chi^2_S$ | Within |
| White Men | $\underline{T}'\underline{T}$ | $\chi^2_T$ | Each |
| Black Women | $\underline{U}'\underline{U}$ | $\chi^2_U$ | Group |
| White Women | $\underline{V}'\underline{V}$ | $\chi^2_V$ | |
| Black vs. White | $(\underline{S}+\underline{U})'(\underline{T}+\underline{V})$ | $\mathcal{L}_R$ | Factor 1 |
| Men vs. Women | $(\underline{S}+\underline{T})'(\underline{U}+\underline{V})$ | $\mathcal{L}_S$ | Factor 2 |
| Interaction | $(\underline{S}+\underline{V})'(\underline{T}+\underline{U})$ | $\mathcal{L}_I$ | Interaction |
| TOTAL | $(\underline{S}+\underline{T}+\underline{U}+\underline{V})'(\underline{S}+\underline{T}+\underline{U}+\underline{V})$ | $\chi^2_C$ | Total Combined Group Agreement |