

# A MEASUREMENT OF AGREEMENT AMONG JUDGES FROM DIFFERENT BACKGROUNDS IN ANALYTIC HIERARCHY PROCESS

Indrani Basak  
Penn State Altoona  
Altoona, PA

## ABSTRACT

In Analytic Hierarchy Process (AHP), decision on priorities of alternatives is made based on judgments from a group of individuals. Often these groups are of different backgrounds (social, geographical or gender) and it is of interest to see whether these different backgrounds create any difference in their prioritization of the alternatives in the decision problem. In this article, we develop a categorical data methodology to measure the level of agreement among groups from different backgrounds regarding their priorities on a fixed set of alternatives.

Key Words: Association, Agreement, Backgrounds, Groups, Cohen's Kappa.

## 1 Introduction

When a group of individuals participate in a decision process by providing their judgments, their thinking can be close or far apart. One important question in a decision problem is whether backgrounds (for example, social, geographical, gender) create a significant disagreement. In this article, we propose a measurement of agreement among groups of judges from different backgrounds using the over-all priority weights of the alternatives for these groups of judges.

## 2 Literature Review

There is a body of statistical literature discussing research designs of the measurement of agreement of the judges based on the standard Analysis of Variance (ANOVA) models for which the measurement scale is assumed to be quantitative. However, standard ANOVA procedures are rarely appropriate for categorical data. Cohen (1960) introduced the concept of the coefficient of agreement for categorical data. He distinguishes the concept of "agreement" from usually determined concept of "association" for categorical data and introduced the coefficient of agreement as the Kappa statistic  $\kappa = \frac{p_0 - p_e}{1 - p_e}$  in which  $p_0$  is the proportion of units in which the judges agreed and  $p_e$  is the proportion of units for which the agreement is expected by chance. Landis (1977) used a weighted Kappa statistic to measure agreement.

## 3 Hypotheses/Objectives

In this article, we propose a measurement of agreement between judges coming from  $B$  many backgrounds regarding their preferences on  $t$  alternatives in the decision problem of AHP. Response variable is the

possible rank order of preference of  $t$  alternatives.

## 4 Research Design/Methodology

Let  $r = L^B$  response profiles be indexed by a response profile  $\mathbf{j} = (j_1, j_2, \dots, j_B)$  where  $j_b = 1, 2, \dots, L$  for  $b = 1, 2, \dots, B$ . Now let  $\pi_{\mathbf{j}} = \pi_{j_1, j_2, \dots, j_B}$  represent the joint probability of response profile  $\mathbf{j}$  for randomly selected groups from  $B$  backgrounds. Then let the marginal probability  $\phi_{bk} = \sum \dots \sum_{\mathbf{j} \text{ with } j_b=k} \pi_{j_1, j_2, \dots, j_B}$  represents the probability of the  $k$ th response category for the  $b$ -th background, where  $b = 1, 2, \dots, B; k = 1, 2, \dots, L$ . Then the weighted  $\kappa$  measure of agreement corresponding to  $h$ th ( $h = 1, 2, \dots, u$ ) set of weights can be formulated by the weighted Kappa statistic  $\kappa_h = \frac{\lambda_h - \gamma_h}{1 - \gamma_h}$  where  $\lambda_h = \sum \dots \sum_{\mathbf{j}} w_{h\mathbf{j}} \pi_{\mathbf{j}}$  and  $\gamma_h = \sum \dots \sum_{\mathbf{j}} w_{h\mathbf{j}} \pi_{\mathbf{j}}^{(e)}$  with  $\pi_{\mathbf{j}}^{(e)} = \pi_{j_1, j_2, \dots, j_B}^{(e)} = \prod_{b=1}^B \phi_{bj_b}$  being the expected probability under the hypothesis of no association among the  $B$  backgrounds. An alternative approach to the above formulation involving all  $B$  backgrounds is to consider pairwise agreement between backgrounds using a standard multiple comparison procedure in statistical theory.

## 5 Data/Model Analysis

Ranges of  $\kappa_h$  “< 0.00”, “0.00-0.20”, “0.21-0.40”, “0.41-0.60”, “0.61-0.80” and “0.81-1.00” are used to measure the relative strength of agreement as “Poor”, “Slight”, “Fair”, “Moderate”, “Substantial” and “Almost Perfect” respectively. Although these divisions are clearly arbitrary, they do provide useful “benchmarks” for measuring the relative strength of agreement and are used in practice.

## 6 Limitations

When the number of alternatives is large in a decision problem, the number of groups from each background need to be also large to avoid missing entries of the multi-dimensional contingency table.

## 7 Conclusions

Judgments and rank orderings are fundamentally based on categorical data and so standard statistical procedures based on numerical methods may not be the best way of handling these types of decision problem. In this article, we propose a measurement of agreement among groups of different backgrounds to check whether or not a set of common priority weights can be established for the combined group of those judges based on a categorical data methodology.

## 8 Key References

- Basak, I. (1988). When to Combine Group Judgments and When Not to in the Analytic Hierarchy Process: A New Method. *Mathematical Computational Modeling*, Vol. 10, No. 6, 395-404.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. XX, No. 1, 37-46.
- Landis, J. R. and Koch, G. G. (1960). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, Vol. 33, 159-174.