Reprinted from IFPress (1999) 2(4)2,3

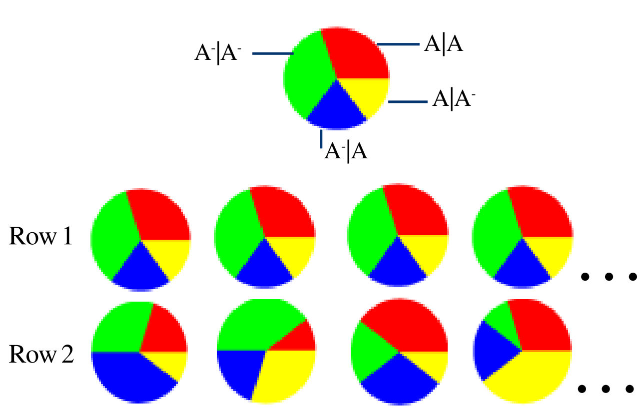
Models for the A-Not A Method

Jian Bi and Daniel M. Ennis

Background: The A-Not A method is a well-known discrimination method used in sensory analysis and signal detection[1]. The method typically involves either paired or sequential presentation of two products; on each trial the subject decides whether the product presented is A or Not A (A^-). A is a product familiar to the subject, sometimes called the "signal" in signal detection experiments. The A-Not A method has been used extensively to study sensitivity and bias in individual subjects and can be used with trained or expert panels. It also has important, though less well known, applications in product testing involving consumer populations where the consumer has developed familiarity with a product through continuous use. Most product testing methods currently in use do not exploit this consumer familiarity or expertise. A potential difficulty in using the A-Not A method is that individual differences may introduce extra variation that cannot be accounted for using the conventional chi-square test commonly used for the analysis of data from this method. If this extra variation is not accounted for, erroneous significance levels will be reported. In this article we discuss an adjustment to the test statistic for this method that accounts for the effect of individual differences.

Individual Differences: Figure 1 illustrates an average outcome for each of four possibilities - $A|A$, $A|A^-$, $A^-|A$ and $A^-|A^-$. In each case the first symbol refers to the response and the second to the product. Below the average result, there are two rows of pie charts. In the first row, the probability of each of the four outcomes is the same as the average. This result would occur if each person tested was identical. The second row of pie charts shows results originating from possibly different individuals. Individual differences are not accounted for in conventional chi-square tests, so these tests tacitly assume that subject populations are composed of clones, at least with respect to choice behavior.

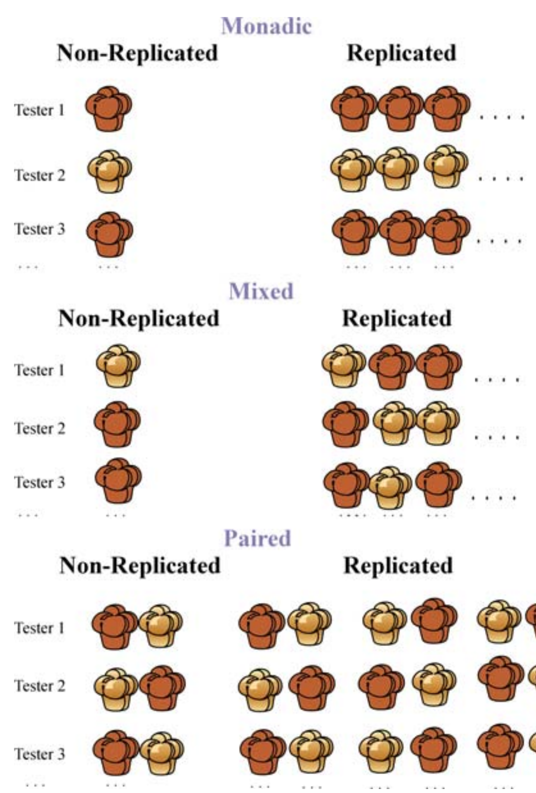
Figure 1. An average set of probabilities for a population of clones (Row 1) or possibly different individuals (Row 2).



The conventional analysis of the A-Not A method involves Pearson's chi-square, irrespective of the design. This approach not only fails to consider individual differences, but also does not recognize differences among variants of the A-Not A method.

Variants of the A-Not A Method: In a recent paper[2], we discussed six variants of the A-Not A method. There are three types of design, and each may be replicated. In the monadic design, each consumer receives only one type of product, either A or A^-. In the mixed design, the consumer may receive both types of products over repeated trials. Moreover, the experimenter does not know in advance exactly how many A or A^- samples will be tested. In the paired design, each consumer receives both A and A^-. These designs are illustrated in Figure 2. Although the non-replicated monadic and mixed designs seem identical, their null hypotheses are different. Table 1 shows how the methods can be categorized, and provides their corresponding models. In the monadic design the test involves a comparison of proportions and is referred to as an homogeneity test. In the mixed design the purpose of the test is to determine whether the response is independent of the sample. This is referred to as an independence test.

Figure 2. Illustration of the monadic, mixed and paired designs.



Non-replicated tests are incapable of detecting individual differences. When replicated designs are used, however, the effect of individual differences can be accounted for. In order to account for variation due to individuals, it is necessary to assume that individual response patterns follow a specified distribution. When there is an observation proportion, as is the case in the monadic and paired designs, the beta distribution is used. When there is an observation vector, as is the case in the mixed design, the Dirichlet distribution is used.

Table 1. Models for variants of the A-Not A method. BB and DM mean beta-binomial[3] and Dirichlet-multinomial[4] respectively.

	Non-Replicated	Replicated
Monadic	χ^2_p Pearson chi-square statistic for homogeneity test	$\tilde{\chi}^2_p$ Adjusted Pearson statistic based on the BB model
Mixed	χ^2_p Pearson chi-square statistic for independence test	$\tilde{\tilde{\chi}}^2_p$ Adjusted Pearson statistic based on the DM model
Paired	χ^2_m McNemar chi-square statistic for correlated proportion test	$\tilde{\chi}^2_m$ Adjusted McNemar statistic based on BB model with unequal replications

Scenario: A modification to a commercial baking process may cause bread rolls to have a greater burned taste than the conventional process. You have access to an experienced group of 20 panelists who have extensive experience in testing daily production of bread rolls and can identify burned taste. In order to test if there is a detectable difference between the conventional and new processes, you decide to use the A-Not A method. In particular you use the replicated mixed design. The rolls baked with the new process are referred to as "A" while the rolls baked with the conventional process are referred to as "Not A" (A^-).

Design: A sample pool composed of A and A^- samples in a ratio of 1:1 is prepared. Each panelist will evaluate 10 samples drawn randomly from this pool. Two hundred total samples will be evaluated in the experiment, but it is not known in advance how many A and how many A^- samples will be evaluated.

Results: You run the experiment and obtain the results summarized in Table 2. Ninety-seven A samples were presented of which 54 were called "A". One hundred and three A^- samples were presented of which 61 were called A^-.

Table 2. Results for replicated mixed A-Not A method.

		Sample A	Sample A^-	Total
Response	A	54	42	
	A^-	43	61	
Total				200

Pearson's Chi-square Model: This model assumes that the 200 responses are independent of each another, and uses Pearson's chi-square to test for independence. This model also assumes that the subjects are identical. In our example, Pearson's chi-square statistic is 4.439 with one degree of freedom; the associated *p*-value is 0.035. Based on this analysis, you would conclude that the new baking process produces rolls that differ in burned taste from the conventional process, but you would have ignored differences among individuals.

Adjusted Pearson's Chi-square Model: Since there are replications for each subject, this model can account for differences among panelists. The effect of individual differences is included by adjusting Pearson's chi-square using the Dirichlet-multinomial (DM) distribution[5]. The adjusted chi-square has one degree of freedom, and follows a chi-square distribution. In order to make the adjustment, it is necessary to calculate an inflation coefficient, C, that measures the extent of individual differences. You find this value to be 2.5, and that this value is significantly greater than 1. The adjusted chi-square value is 1.776 (4.439/2.5); its associated *p*-value is 0.183. You now conclude that there is no evidence at the 5% level that the new process is different from the conventional process with respect to burned taste. The apparent difference found earlier may have been due to individual differences. The size of the C value suggests that the panel chosen was not homogenous. In order to improve power it would be useful to study and improve the sensitivity of the panelists. Panels can be evaluated for homogeneity using the C statistic and the effects of training and selection can thus be monitored.

Conclusion: The A-Not A method can be of great use in product testing particularly with experienced panels. We discussed six designs for this method and showed that there is a corresponding model associated with each design. The main point of this report is that consumer populations and product testing panels exhibit individual differences. The effect of these differences must be taken into account when analyzing data from the A-Not A method. The importance of this recommendation was illustrated using data from the replicated mixed A-Not A design.

References:

1. International Standard ISO 8588 (1987). Sensory analysis-Methodology-"A"-"Not A" test.
2. Bi, J. and Ennis, D. M. (1999). Statistical models for the A-Not A method. Submitted.
3. Ennis, D. M. and Bi, J. (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies,* **13**, 461-466.
4. Ennis, D. M. and Bi, J. (1999). The Dirichlet-multinomial model: Accounting for inter-trial variation in replicated ratings. *Journal of Sensory Studies,* **14**, 321-345.
5. Brier, S. S. (1980). Analysis of contingency tables under cluster sampling. *Biometrika,* **67**, 591-596.