Food Quality and Preference 32 (2014) 48–55

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Paired preference data with a no-preference option – Statistical tests for comparison with placebo data

Rune Haubo Bojesen Christensen [a,*], John M. Ennis [b], Daniel M. Ennis [b], Per Bruun Brockhoff [a]

[a] DTU Compute, Statistics Section, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kongens Lyngby, Denmark
[b] The Institute for Perception, Richmond, VA, USA

ARTICLE INFO

Article history:
Received 28 September 2012
Received in revised form 17 May 2013
Accepted 1 June 2013
Available online 18 June 2013

Keywords:
2-AC
Paired preference
Type I error
Power
Simulation study

ABSTRACT

It is well-established that when respondents are presented with identical samples in a preference test with a *no preference* option, a sizable proportion of respondents will report a preference. In a recent paper (Ennis, D. M., & Ennis, J. M. (2012a). Accounting for no difference/preference responses or ties in choice experiments. *Food Quality and Preference, 23*, 13–17) noted that this proportion can depend on the product category, have proposed that the expected proportion of preference responses within a given category be called an *identicality norm*, and have argued that knowledge of such norms is valuable for more complete interpretation of 2-Alternative Choice (2-AC) data. For instance, these norms can be used to indicate consumer segmentation even with non-replicated data. In this paper, we show that the statistical test suggested by Ennis and Ennis (2012a) behaves poorly and has too high a type I error rate if the identicality norm is not estimated from a very large sample size. We then compare five χ^2 tests of paired preference data with a *no preference* option in terms of type I error and power in a series of scenarios. In particular, we identify two tests that are well behaved for sample sizes typical of recent research and have high statistical power. One of these tests has the advantage that it can be decomposed for more insightful analyses in a fashion similar to that of ANOVA *F*-tests. The benefits are important because they enable more informed business decisions, particularly when ingredient changes are considered for cost-reduction or health initiative purposes.

1. Introduction

Difference testing is as relevant as ever given the numerous ingredient change projects currently underway for cost-reduction or health-initiative purposes. Thus, there is presently much interest in interpreting difference testing results in as reliable, as meaningful, and as powerful a manner as possible (Bi, Lee, & O'Mahony, 2011; Brockhoff & Christensen, 2010; Christensen & Brockhoff, 2009; Christensen, Cleaver, & Brockhoff, 2011; Christensen, Lee, & Brockhoff, 2012; Ennis & Ennis, 2012b; Ennis & Jesionka, 2011; Hautus, Shepard, & Peng, 2011; Ishii, Kawaguchi, O'Mahony, & Rousseau, 2007; Lee, van Hout, & Hautus, 2007; van Hout, Hautus, & Lee, 2011).

In a recent paper, Ennis and Ennis (2012a) developed the idea of an identicality norm for 2 Alternative Choice (2-AC) data. The 2-AC protocol is a 2-AFC protocol with a *no difference* option and it is from a mathematical viewpoint equivalent to the paired preference test with a *no preference* option.[1] The identicality norm is obtained by conducting a paired preference test with a *no preference* option with identical products – the idea is similar to that of using a placebo drug in a medical trial. The expected distribution over "Prefer A", "No Preference" and "Prefer B" for the identical products then constitutes the identicality norm.

The identicality norm can be useful in a situation where a paired preference test with a *no preference* option has been conducted, but the products appear to be approximately equally preferred and a conventional statistical test, e.g. a Pearson test does not show significant differences in preference between the two products in question. However, the products might appear equally preferred if the consumer sample consists of two segments with opposite preferences; in this case preferences may approximately balance out in the sample and the products appear to be equally preferred while in fact most consumers have a preference for either of the products.

Ennis and Ennis (2012a) observed that the data table can be compared to an identicality norm in a Pearson χ^2 test, and that this test can indicate if opposing segments of preference balance out over the sample as just described. But, as we will see in this paper, the statistic proposed by Ennis and Ennis (2012a) only performs well if the identicality norm is based on a placebo experiment with a very large sample size. Otherwise, as we will show, the failure of

* Corresponding author. Tel.: +45 3026 4554.
E-mail address: rhbc@imm.dtu.dk (R.H.B. Christensen).

[1] For the remainder of this article, we will refer only to *no preference* votes, for simplicity.

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.06.005

this statistic to take into account the variability within the placebo experiment leads to an unacceptably large type I error.

The purpose of this paper is to advance the statistical analysis of 2-AC data with placebo experiments by comparison of five statistical tests. To this end we propose two tests that are well behaved for sample sizes typical of recent research (Alfaro-Rodriguez, Angulo, & O'Mahony, 2007; Chapman & Lawless, 2005; Kim, Lee, O'Mahony, & Kim, 2008; Marchisano et al., 2003) and have high statistical power. One test has the feature that asymptotically, as the sample size for the placebo experiment approaches infinity, the χ^2 test suggested by Ennis and Ennis (2012a) is obtained. The other test has the advantage that it may be conveniently decomposed into directional and tie effects in a fashion similar to that of ANOVA *F*-tests.

Throughout the paper we assume that there is no difference between AA and BB placebo pairs. In a Thurstonian view, the AA and BB placebo pairs could be expected to distribute differently if their perceptual variances were different. It is still an open question to what extent the tests discussed in this paper are valid if the distributions of AA and BB placebo pairs differ. If the design is based on an approximate equal number of AA and BB placebo pairs, the tests are probably still appropriate.

In Section 2 example data from Ennis and Ennis (2012a) are reanalyzed illustrating that the uncertainty in the placebo experiment is not taken into account. In Section 3 five test statistics are presented, problems with the genuine Pearson test are exposed and alternative tests are suggested. In Section 4 these five tests are compared in terms of type I error rate and power in a series of scenarios. In Section 5 we end with discussion and recommendations. All computations were done in **R** (R Development Core Team, 2011) and the code to perform all simulations are available in the online Supplements.

2. χ^2 tests with identicality norms

To motivate the adjustments to the χ^2 test that was originally suggested by Ennis and Ennis (2012a), we will use the example presented in section 4 in Ennis and Ennis (2012a). In this example it is assumed that the following triplet of data have been obtained (25, 15, 60 for "Prefer A", "No Preference", "Prefer B"), and that the identicality norm can be assumed to be 40%, 20% and 40% for those three response options.

Ennis and Ennis, 2012a in essence suggest that we compute expected values as $100 \cdot (0.4, 0.2, 0.4) = (40; 20, 40)$ and compare these to the observed values, $(25, 15, 60)$ in a Pearson χ^2 test. The test statistic is

$$X^2 = (25-40)^2/40 + (15-20)^2/20 + (60-40)^2/40 = 5.625 + 1.250 + 10.00 = 16.875.$$

Comparing this value to a χ^2_2 distribution (with 2 degrees of freedom) yields a *p*-value of 0.00022 as also found by Ennis and Ennis (2012a).

Observe that this test does not depend on the sample size involved in setting the identicality norm, hence the identicality norm is inherently assumed to be known without error. If the identicality norm is determined without any uncertainty all is well, but if an experiment with identical products as described in the introduction was used to obtain the identicality norm, it will be determined with some uncertainty, and it is desirable to take account of that uncertainty in the statistical test.

Intuitively we expect that if the identicality norm is obtained using a large sample size, it is accurately determined and the results should not change much. If, on the other hand, the identicality norm is obtained from a small sample size, the norm is more uncertain and it should be harder to get a significant result.

Table 1
Observed counts for example in Section 2.

	"Prefer A"	"No Preference"	"Prefer B"	Total
Preference experiment	25	15	60	100
Placebo experiment	40	20	40	100
Total	65	35	100	200

Table 2
Expected values for the observed counts in Table 1.

	"Prefer A"	"No Preference"	"Prefer B"	Total
Preference experiment	32.5	17.5	50.0	100
Placebo experiment	32.5	17.5	50.0	100
Total	65.0	35.0	100.0	200

Now assume, for instance, that the identicality norm in our example was determined from a placebo experiment with 100 observations. We can then arrange the data in the 2×3 table shown in Table 1. The corresponding table of expected values are given in Table 2. As an example, the expected value in the $(1, 1)$ cell is obtained as $100 \cdot 65/200 = 32.5$ since the sum in the first row is 100, the sum in the first column is 65 and the total sum is 200. Computing the Pearson χ^2 test on these tables now yields

$$X^2 = \frac{(25-32.5)^2}{32.5} + \frac{(40-32.5)^2}{32.5} + \cdots + \frac{(40-50.0)^2}{50.0} = 8.18$$

The number of degrees of freedom are $(2-1)\cdot(3-1) = 2$, and in comparison with the χ^2 distribution we obtain a *p*-value of 0.0168. The Pearson χ^2 test for association in this table is a test for whether the paired preference test data are in compliance with the placebo data, therefore essentially the same hypotheses are tested as in the test suggested by Ennis and Ennis (2012a). This *p*-value is larger reflecting that the uncertainty in the identicality norm is taken into account. While the sample size for the identicality norm is still large enough that the test is significant, this will change if a smaller sample size for the placebo experiment is assumed.

Table 3 shows the value of the Pearson χ^2 statistic and *p*-value for a range of sample sizes used to determine the identicality norm. The table shows that the smaller the sample size used for setting the identicality norm, the larger the *p*-value. Had, for instance, the sample size for the identicality norm been 50, the test would not have been significant on the 5% limit. Table 3 also shows that the test suggested by Ennis and Ennis (2012a) is obtained in the limit as the sample size used to determine the identicality norm

Table 3
Pearson χ^2 statistic and *p*-value for a range of sample sizes for the placebo experiment.

n	Statistic	*p*-Value
10	1.55	0.46118
20	2.80	0.24619
30	3.85	0.14601
40	4.74	0.09371
50	5.50	0.06393
60	6.17	0.04578
70	6.76	0.03410
80	7.28	0.02624
90	7.75	0.02074
100	8.18	0.01677
1000	15.15	0.00051
10^4	16.68	0.00024
10^5	16.86	0.00022
10^9	16.87	0.00022