ON THE NUMBER OF PERCEIVERS IN A TRIANGLE TEST WITH REPLICATIONS

Michael Meyners

Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany
E-mail: michael.meyners@udo.edu
Fax: +49 (0)231 755 3454

ABSTRACT

In discrimination tests, two different questions usually arise: First of all, we are interested in deciding whether or not there are product differences at all that might be perceived by the assessors. However, often this is not our most important concern, since the main question is whether or not the consumers (in contradiction to e. g. a trained panel) might perceive the difference and, if so, how many of them are supposed to do so. While the first question has been addressed frequently in recent times, the known models for estimating the proportion of perceivers use strong conditions, e. g. that the assessors taste the difference always or never. We propose a more general model that allows the assessors to perceive differences once in a while and derive a method that takes this assumption into account. Several examples show that the estimates for the proportion of interest are quite reasonable.

Key-words : triangle test, replications, perceivers, multiple test theory

FRENCH ABSTRACT

Il existe deux questions d'intérêt autour des tests de discrimination: Il s'agit d'abord de savoir s'il existe des différences entre les produits perceptibles par les sujets ayant participé au test. Mais le plus important est souvent de décider si les consommateurs (et non pas des sujets entraînés) perçoivent la différence entre les produits et, plus précisément, quelle proportion de la population la perçoive? Différentes contributions récentes traitent de la première question, mais les modèles pour estimer la proportion des consommateurs percevant la différence sont fondés sur des hypothèses restrictives, par exemple qu'un sujet qui perçoit la différence réussira toutes les répétitions du test de différence. Nous proposons un modèle et une méthode plus générales qui prennent en compte que les consommateurs peuvent ne percevoir la différence que de temps en temps. Nous présentons plusieurs exemples qui montrent que cette méthode conduit à des résultats logiques.

Mots clés : test triangulaire, répétitions, sujets discriminants, inférence statistique multiple

1. Introduction

Consider the triangle test with n assessors and k replicates. Kunert (2001) proposes a model to estimate the proportion of consumers that are able to perceive a difference between the products. Furthermore, he calculates a confidence band for this proportion. Anyway, he considers only the worst case in which each assessor perceives the difference always or never, thus giving a success probability of either $\frac{1}{3}$ or 1. Even though this is a useful first approach, we think that due to variations within a product or fatigue effects, differences might occur to the assessors only during some of the replicates. Furthermore, if some of the assessors can always figure out the difference, these should be so apparent that we would expect most of the other assessors to be able to perceive the difference at least once in a while. As well as Brockhoff and Schlich's (1998) does, our model takes into account that the replicates of a perceiver are not independent. Anyway, we think we should get more knowledge from the information how often each assessor figured out the right answer. On the other hand we make only weak distribution assumptions on the random variables. Of course we consider the simple binomial test given by Kunert and Meyners (1999) the right method to examine whether or not there are detectable product differences at all, irrespective of whether or not we consider replicates. Nevertheless we think that the sensory analyst usually is rather interested in estimating the proportion of perceivers than in deciding whether or not there are differences at all. Thus we propose the use of a different approach that deals with more realistic circumstances. Assume the following artificial example to illustrate the difference: Consider $n = 20$ and $k = 3$. In a first trial, ten persons give one correct answer each (which is exactly what we expect under the null hypothesis of product equality) and another ten giving three right answers each, i. e. they always identify the odd sample. In all we have 40 right answers in 60 replicates, so an estimate of the proportion of perceiving assessors assuming them to have success probability either $\frac{1}{3}$ or 1 would be $\left(\frac{40}{60} - \frac{1}{3}\right) \cdot \frac{3}{2} = \frac{1}{2}$, i. e. a half of the consumers is judged to be able to perceive the difference. Fortunately, this seems to be a reasonable estimate for this data. But now let us assume the case in which all 20 assessors gave two right answers. Using the same estimate we would also estimate a half of the consumers to perceive the difference. Despite the fact that in case the model holds and there are any perceivers participating the test we might not have observed this result (since from the model they are thought to find the odd sample in every replicate), we think that there is a much larger portion of assessors that really detected a difference at least once or twice. If we had only *three* assessors with two right answers each, using the method of Kunert and Meyners (1999) we would also have claimed significant differences between the products, since the probability of observing this result by chance if there are no differences at all is less than 0.05. Thus we might conclude that at least 18 persons must have been able to perceive the difference once or twice to obtain this result. Hence we would estimate one non-perceiver in ten in comparison to one in two as before, which gives quite a different conclusion for the analyst.

2. Model assumptions

Let δ denote the portion of perceivers within the population of interest and π_0 the probability to succeed by chance, i. e. if an assessor tastes no differences between the samples. For the triangle test we have

$$\pi_0 = \frac{1}{3},$$

whereas for other discrimination tests like e. g. the duo-trio-test, we would have to consider a different value of π_0. Let $\varepsilon \in (0,1)$ denote the minimal success probability of interest, e. g. if we are interested in consumers detecting the difference once in ten times, we have $\varepsilon = 0.1$. Furthermore for $i = 1,..., n$ let η_i denote non-negative random variables with values in $(0, 1 - \varepsilon)$. The values of the variables might vary between assessors i as well as their distribution may and will not be restricted furthermore. We will neglect this term later on for estimation purposes using a worst case scenario, anyway, it allows to treat a much larger class of possible models with the same method. Then we assume the success probability of a perceiver to be

$$\frac{1}{3} + \frac{2}{3}(\varepsilon + \eta_i).$$

The restrictions on ε and η_i guarantee that this probability does not exceed the natural boundaries $\frac{1}{3}$ or 1. Then for an assessor who has been randomly drawn from the population of interest, his / her success probability is given by

$$(1-\delta)\frac{1}{3} + \delta\left[\frac{1}{3} + \left(1 - \frac{1}{3}\right)(\varepsilon + \pi_i)\right] = \frac{1}{3} + \frac{2}{3}\delta\ (\varepsilon + \pi_i).$$

3. Estimating δ

To estimate the value of δ we use the following procedure: We start with the usual binomial test proposed by Kunert and Meyners (1999), i. e. we consider all k replicates of all n assessors to be independent and test the hypotheses

$$H_0: \ \pi = \frac{1}{3}$$

versus

$$H_1: \ \pi > \frac{1}{3}$$

using the binomial distribution with $n \cdot k$ observations and probability parameter $\frac{1}{3}$ at a significance level α, say. If we cannot reject the nullhypothesis we stop and conclude that we cannot prove any differences between the products to be apparent for the consumers, at least not with these assessors. With it, of course we estimate δ to be zero, i. e. $\hat{\delta} = 0$. Otherwise if we find a significant difference, we reduce the data set removing the assessor with the most correct answers respectively one of them when there are several with the same number of successes. (Note that we do not investigate on *which* assessors are perceivers, thus without loss of generality we can cross out any of those.) The assessor crossed out is considered to be a perceiver and we recalculate the binomial test with the results of the remaining $n - 1$ assessors, using the binomial distribution with $(n - 1) \cdot k$ observations and parameter $\frac{1}{3}$. If this test gives no significance, we stop and estimate a portion $\hat{\delta} = \frac{1}{n}$ of the consumers to be perceivers. Else if significance is found, we cross out another assessor claiming him or her to be a perceiver, and we repeat the binomial test with the results of the remaining $n - 2$ assessors. We go on with this procedure until non-significance occurs for the first time. The number of assessors that have been crossed out, r, say, is used to estimate δ by

$$\hat{\delta} = \frac{r}{n}.$$

A theoretical justification of this approach using the multiple test theory will be given elsewhere, instead of this we reconsider the artificial data given in the introduction to get a first impression of the performance of this method. Choosing $\alpha = 0.05$, in the first case with ten assessors giving three right answers each and another ten with one success each, we need nine iterations to find non-significance, i. e. we crossed out eight assessors. Thus we estimate the proportion of interest to be

$$\hat{\delta} = \frac{8}{20} = 0.4\,,$$

which is not too far away from the estimate of the naïve approach giving $\frac{1}{2}$. Anyway, in case of 20 assessors with two right answers each, the naïve approach again gives a value of $\frac{1}{2}$, whereas our method needs 19 iterations removing 18 assessors, hence we have

$$\hat{\delta} = \frac{18}{20} = 0.9\,.$$

This is far away from both the first case and the naïve approach. However, we think that this estimate represents the given structure of the data much better, since there are strong hints within that there have been quite a lot more than seven or eight assessors perceiving a difference at least once or twice. Hence if we are interested not only in those assessors identifying the difference in each replicate, we should estimate the portion of perceivers the way proposed here.

4. Examples

We re-analyze some data given in the literature to compare the results of our approach with those given elsewhere. To start with we use the three data sets presented by Hunter, Piggott and Monica-Lee (2000) which where also analyzed by Kunert (2001). In all trials the assessors were asked to carry out the test k = 12-times, whereas n varied from 30 in the first over 24 in the second to 23 in the last set. Table 1 represents the number of assessors that gave x right answers, $x \in \{0,1,...,12\}$. Using a level of 5%, for the first data set our method identifies 13 perceivers, thus leading to

$$\hat{\delta} = \frac{13}{30} \approx 0.43 .$$

Hence we conclude that at least 2 of 5 consumers figure out the difference once in a while. Using the naïve approach, δ would be estimated by

$$\hat{\delta} = \left(\frac{170}{360} - \frac{1}{3} \right) \cdot \frac{3}{2} \approx 0.21 ,$$

which is much smaller and might lead to a quite different interpretation.

x	0	1	2	3	4	5	6	7	8	9	10	11	12	total number of successes
experiment 1	0	0	1	2	3	7	8	6	2	1	0	0	0	170
experiment 2	1	0	1	5	5	3	3	3	1	2	0	0	0	117
experiment 3	0	0	2	1	1	4	3	6	3	1	0	1	1	147

Table 1: Number of assessors with x right answers for the three experiments reported by Hunter *et. al.* (2000).

For the second experiment we get a quite different result since we find only two assessors that have to be assumed to be a perceiver. Hence we calculate

$$\hat{\delta} = \frac{2}{24} \approx 0.08 ,$$

which is very small and quite similar to the naïve value which is about 0.11.

However, at least we can prove significant differences between the products which is not achieved using the method of Brockhoff and Schlich (1998).
Using our method, the number of perceivers in the third experiment is estimated to be 11 from 23. Hence we get

$$\hat{\delta} = \frac{11}{23} \approx 0.48\,.$$

In comparison, using the naïve approach we estimate δ to be 0.30.

Finally we consider the data set given in the first example of Brockhoff and Schlich (1998) and which can be found in table 2. This data contains the results for n = 12 assessors and k = 4 trials each. In this experiment, using our approach we find only one perceiver out of 12 assessors, thus $\hat{\delta} \approx 0.08$, which this time is smaller than the naïve estimator of about 0.25. Anyway, again we can at least prove differences between the products.

x	0	1	2	3	4	total number of successes
number of assessors	2	2	4	2	2	24

Table 2: Number of assessors with x right answers for the first experiment reported by Brockhoff and Schlich (1998).

We might also estimate a confidence interval for δ calculating the upper and lower limits for π, π_L and π_U, say, by searching the values for which we would not have observed a significant result using a test for

$$H_0: \ \pi > \pi_L$$

respectively

$$H_0: \ \pi < \pi_U$$

against the appropriate alternatives. However, these intervals strongly depend on the parameter ε which gives us the relevant proportion of replicates in which a consumer tastes a difference. In the first example of Hunter *et. al.* (2000), these intervals vary from [0.575,1] over [0.345,0.885] to [0.230, 0.590] for reasonable values of ε, so they do not even necessarily include the estimators. Thus the value of ε is to be carefully chosen. The details of the estimation of confidence bands are beyond the scope of this paper.

5. Conclusion

We have proposed an alternative approach to estimate the number of perceivers in a triangle test with replications. For different examples we have shown that this approach gives reasonable estimates for the proportion of interest. Furthermore, even though not explicitly shown in details here, a theoretical justification of this approach can be found using the theory of multiple tests while the estimates are still quite easy to calculate. Hence this procedure might be used whenever a triangle test is considered. If we have no replicates, of course we might use the naïve approach as well, but whenever there are replications, the structure of the data can be represented more reasonably by our method. To calculate appropriate confidence intervals, we have to find a reasonable value of ε, indicating the proportion of successful replicates about which the analyst might worry in applications. Finally it has to be stated that, of course, the results can be easily adapted to other discrimination tests.

Acknowledgements

The author is grateful to the Deutsche Forschungsgemeinschaft (SFB 475, “Reduction of complexity for multivariate data structures”) for the financial support of this work.

References

Brockhoff P.B. Schlich P.	1998	*Handling replications in discrimination tests.* Food Quality and Preference 9, 303 – 312.
Kunert J.	2001	*On repeated difference testing.* Food Quality and Preference 12, 385 – 391.
Kunert J. Meyners M.	1999	*On the triangle test with replications.* Food Quality and Preference 10, 477 – 482.
Hunter E.A. Piggott J.R. Monica-Lee K.Y.	2000	*Analysis of discrimination tests.* Agro-industrie et methodes statistiques, Pau, january 19-21,2000.