Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture 1995 - 7th Annual Conference Proceedings

SEQUENTIAL ANALYSIS OF AGRICULTURAL EXPERIMENTS

Armando Garsd

María V. López

María del C. Fabrizio

Follow this and additional works at: https://newprairiepress.org/agstatconference

 Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Garsd, Armando; López, María V.; and Fabrizio, María del C. (1995). "SEQUENTIAL ANALYSIS OF AGRICULTURAL EXPERIMENTS," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1339

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

SEQUENTIAL ANALYSIS OF AGRICULTURAL EXPERIMENTS

Garsd, Armando, López, María V., and Fabrizio, María del C.
Dep. of Statistics, School of Agriculture, University of Buenos Aires

Abstract

Interim monitoring of accumulating data has been widely used in clinical trials, but it has not received the same attention in agricultural experimentation. The methodology, however, can be a useful tool in agronomic trials designed to find better production techniques or optimal animal treatments at low cost, plus the possible economic advantages resulting from correct early decisions. These sequential procedures for testing hypothesis with available data in successive periods of time dictate termination of the experiment when a significant difference is detected, or otherwise continuation of the experiment to the end of the stipulated time or until all the planned sample size is realized. The statistical cost of repeated testing of part of the same data is a reduction in the significance levels α to the time-related significance levels α_j ($\alpha_j<\alpha$). We apply three methods for this type of analysis, which we illustrate with two examples involving respectively, comparisons of two proportions and two means from normally distributed random variables with unknown variances. The examples show the usefulness and limitations of the proposed methods and also that there can be no absolute rule for choosing the best method of analysis in a particular case. The optimal strategy depends on the specifics of the trial and the investigator's criterion to choose the α_j.

Keywords: Stopping rules, interim analysis, agricultural experimentation, sequential testing.

1. Introduction

It is common practice that accumulating data be reviewed periodically during the course of an experiment. In contrast to a single stage test, it is recognized that when analyses are performed repeatedly, some adjustment has to be made to maintain the probability of type I error (α) at a specified level. Armitage, McPherson and Rowe (1969) showed, for example, that testing accumulating data on three successive occasions, each time at $\alpha=0.05$, amounts to working at a 0.11 overall probability of type I error.

Statisticians have proposed various methods to address this issue.

We investigate three of these methods in the context of two examples from agronomy. We restrict attention to the case of two treatments and the number of successive tests fixed a priori.

2. Methods

Consider an experiment to compare the mean responses, μ_A and μ_B, of two treatments, A and B. Suppose that experimental units are entered sequentially and randomized so that each consecutive group of experimental units has n of each treatment.

In the following we use standard notation. The usual hypothesis set-up is H_0: $\mu_A=\mu_B$ vs H_1: $\mu_A\neq\mu_B$. H_0 is tested at the α significance level via

$$z=(\overline{x_A}-\overline{x_B})/\sqrt{(2\sigma^2/n)}$$

or

$$z=(p_A-p_B)/\sqrt{(2\bar{p}(1-\bar{p})/n} \quad ,$$

where $\bar{x}$, p and $\bar{p}$ denote average, proportion and average proportion, respectively, and z is distributed as a standard normal random variable.

Let

$$z_j^*=\overline{d_j}/\sqrt{2\sigma^2/jn}, \quad j=1,2,\ldots,K, \qquad (1)$$

where

$$\overline{d_j}=(\sum_{i=1}^{j}(\overline{x_{Ai}}-\overline{x_{Bi}}))/j \quad or \quad \overline{d_j}=(\sum_{i=1}^{j}(p_{Ai}-p_{Bi}))/j, \qquad (2)$$

K is the total number of interim analyses and α_j is the fraction of type I error allocated to the test performed at time j. Notice that z_j^* involves all the data accumulated up to time j.

In general, the group sequential procedure is

(i) stop and reject H_0 if $|z_j^*| > z_{1-\alpha j/2}$, or

(ii) continue if $|z_j^*| \leq z_{1-\alpha j/2}$, or

(iii) accept H_0 if $|z_K| < z_{1-\alpha K/2}$.

Various functions α_j have been proposed (see, for example, DeMets,1987). Table 1 presents three of these possible functions. Several authors have addressed the issue of how well each of the three methods controls the type I error (Table 2).

If the variance σ^2 is unknown, the random variable z is replaced by (Student) t, based on 2(jn) degrees of freedom. Pocock (1977) showed that if an updated variance has to be estimated for each interim analysis, the resulting overall α is close to the nominal 0.05 and 0.01 levels and that in these cases the loss of statistical power is negligible.

3. Applications

The examples that follow are based on real experiments but for illustrative purposes the results given below were generated from hypothetical populations differing in known parameters. All computations were performed in SAS (SAS System, 1988).

Example 1

A two-year long study was conducted to investigate low power laser beam (A) vs traditional hydrotherapy (B) for the treatment of tendinitis in polo and race horses. Therapy is deemed successful if the horse can compete after two days of treatment. The true underlying proportions were 0.60 and 0.40 for laser and hydrotherapy, respectively.

The standard non-sequential statistical analysis at completion, 2 years later, with 100 horses accrued in each treatment, resulted in 63 and 43 successful A and B treatments, respectively; that is, p_A=0.63, p_B=0.43, z=2.83; p<0.05. The ensuing conclusion is to reject H_0, (i.e.,laser treatment is better).

Consider now a hypohetical retrospective analysis based on five sequential subgroups. Results are presented in Table 3, with Table 4 indicating that by Pocock's criterion the study could have been stopped at j=3, (2.566>2.41), using only 60 horses (i.e., 40 percent less experimental units than the total actually used). However, the other two criteria call for completing the study (j=5), at which point the two criteria also yield significant results in favor of treatment A. It can be seen from this example that Pocock's method, by allocating α evenly to the α_j's, may terminate the study earlier than any of the other two procedures.

Example 2:

Preservative treatment for wooden shingles (De Groot, 1994) manufactured from the Pacific silver fir (*Abies amabilis*). The experimental unit is made of 10 shingles from each tree. Two preservative treatments are tried, A vs B. Due to the high cost of trees, the experiment was performed in three stages of n_j=10 trees per treatment. The outcome variable is the retention of active ingredient reported on a weight/weight percent basis (weight of active ingredient/weight of wood) x 100. Results are given in Table 5. The two means were 1.004% and 0.743% for formulations A and B, respectively. The pooled standard desviation was 0.292%.

It follows from the P_j values in Table 5 and the levels in Table 1 that only Haybittle's method at j=3 attains statistical significance. This example shows that if the data display an increasing trend towards significance, Pocock's method is too conservative, whereas O'Brien-Fleming's method, by being progressively more `lenient', eventually does detect a significant difference, albeit at the end of the study.

4. Summary

Agricultural experiments consume substantial unit, investigator and financial resources. Practical concerns indicate that investigators should not deploy resources inefficiently or unnecessarily. Thus, a study that shows early benefit or unexpected toxicity mandates serious consideration for early termination. Periodic interim statistical analysis on accumulating data are designed to achieve this objective.

However, repeated testing can substantially inflate the type I error rate. It is widely recognized that to control this error some type of adjustment (α_j) has to be made to maintain the probability type I error at the specified level α. We have considered only three α_j functions.

We illustrate the use of these functions with two examples involving, respectively, comparisons of two proportions and two means from normally distributed random variables with unknown variances. The examples show the usefulness and limitations of the proposed methods and also that there can be no absolute rule for choosing the best method of analysis in a particular case. The optimal strategy depends on the specifics of the trial and the investigator's criterion to choose the α_j. The question of which function α_j to use a priori, is a difficult question, and one for which there is no definite answer at this time.

Acknowledgements

We wish to thank Dr. Dámaso Soraires for his assistance with example 1 involving laser therapy in horses.

References

Armitage, P., McPherson, C. K., and Rowe, B. C. 1969. Repeated significance tests on accumulating data. *J. R. Statis. Soc.*, Series A, 132:235-244.

De Groot, R. 1994. Treatability of western softwood and red alder shakes. *Forest Products Journal*, 44: 34-43.

DeMets, D. 1987. Practical aspects in data monitoring: a brief review. *Statistics in Medicine*, 6:753-760.

Haybittle, J. L. 1971. Repeated assessment of results in clinical trials of cancer treatment. *British Journal of Radiology*, 44: 793-797.

Jenninson, C., and Turnbull, B. 1989. Interim analyses: the repeated confidence interval approach. *Royal Statistical Society*, 51: 305-361.

O'Brien, P. C., and Fleming, T. R. 1979. A multiple testing procedure for clinical trials. *Biometrics*, 35: 549-556.

Pocock, S. J. 1977. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64: 191-199.

SAS System. SAS Institute Inc., *SAS/Language guide and SAS/STAT user's guide*, release 6.03 edition, 1988, Cary, NC, USA.

Table 1: Nominal significance levels α_j for two-sided group sequential designs with overall significance level α=0.05

K	Analysis	Pocock (1977)	O'Brien & Fleming (1979)	Haybittle (1971)
2	1	0.029	0.005	0.0027
	2	0.029	0.048	0.050
3	1	0.022	0.0005	0.0027
	2	0.022	0.014	0.0027
	3	0.022	0.045	0.050
4	1	0.018	0.0001	0.0027
	2	0.018	0.004	0.0027
	3	0.018	0.019	0.0027
	4	0.018	0.043	0.050
5	1	0.016	0.00001	0.0027
	2	0.016	0.0013	0.0027
	3	0.016	0.008	0.0027
	4	0.016	0.023	0.0027
	5	0.016	0.041	0.050

k: total number of interim analyses.

Table 2: Control of type I error by sequential methods

Method	Control of global α		Reference
	Distribution		
	Normal	Other	
Pocock	exact	approx.(1)	Pocock (1977)
O'Brien-Fleming	exact	approx.(2)	O'Brien-Fleming (1979) Jenninson-Turnbull (1989)
Haybittle	approx.	approx.(3)	Haybittle (1971)

approx.: Approximately.

(1) Results from simulations based on exponential and binary responses.
(2) Results from simulations based on dichotomous responses.
(3) Results from simulations based on lognormally distributed survival times.

Table 3: Hypothetical retrospective interim analysis based on five subgroups of n_j=20 horses per treatment (Example 1)

n_j	p_{Aj}	p_{Bj}	$\bar{d}j$	$\hat{\sigma}_j$	z_j^*
20	0.60	0.40	0.200	0.1581	1.265
20	0.65	0.40	0.225	0.1118	2.013
20	0.75	0.50	0.233	0.0908	2.566
20	0.50	0.50	0.175	0.0788	2.221
20	0.65	0.35	0.200	0.0706	2.833

p_{Aj}, p_{Bj}: estimated proportions of successes with treatments A and B at j, respectively. $\hat{\sigma}_j$: estimated standard deviation of the differences in proportions at j.

Table 4: Critical limits (z_{cj})

Method	zc1	zc2	zc3	zc4	zc5
Pocock's	2.41	2.41	2.41	2.41	2.41
O'Brien y Fleming's	4.56	3.22	2.63	2.28	2.04
Haybittle's	3.00	3.00	3.00	3.00	1.96

Table 5: Interim analysis for the comparison of preservative treatments for wooden shingles (Example 2)(1)

n_j	$\bar{x}_{Aj}$	$\bar{x}_{Bj}$	$\bar{d}_j$	s_j	t_j*	P_j
10	0.953	0.755	0.198	0.1412	1.412	0.174
10	0.908	0.738	0.184	0.0928	1.982	0.054
10	0.962	0.896	0.145	0.0715	2.023	0.048

(1) measurements are expressed in %.

s_j: estimated standard deviation of the difference in means.
t_j*: value of t-test at time j.
P_j: significance level of the two-sided group sequential test based on the Student t distribution with $2(n_j-1)$ degrees of freedom.