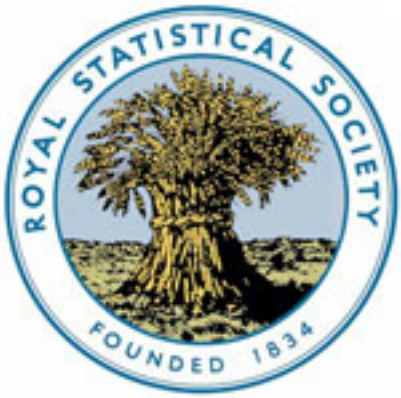

J. R. Statist. Soc. B (1980),
42, *No.* 1, *pp.* 26–30

Selecting the Better Bernoulli Treatment Using a Matched Samples Design

By AJIT C. TAMHANE

Northwestern University, Evanston, Illinois

[Received June 1978. Final revision February 1979]

SUMMARY

The problem of selecting the better Bernoulli treatment using a matched samples design is considered in the framework of the indifference-zone approach. A single-stage procedure is proposed and its properties are studied. Tables of sample sizes for implementing the proposed procedure are given. A comparison is made with an independent samples design and the associated Sobel–Huyett selection procedure.

Keywords: MATCHED SAMPLES DESIGN; INDEPENDENT SAMPLES DESIGN; BERNOULLI TREATMENTS; RANKING AND SELECTION; INDIFFERENCE-ZONE APPROACH

1. INTRODUCTION

IN this paper we consider the problem of selecting the "better" of the two Bernoulli populations (i.e. the one having the larger success probability) when a matched samples design is used. The corresponding problem when the independent samples design is used has been considered by Sobel and Huyett (1957). It should be noted that although considerable literature exists on the problem of comparing matched proportions (see McNemar, 1947; Cochran, 1950; Bennet, 1967, 1968; Bhapkar, 1973; Bhapkar and Somes, 1977), mostly it deals with tests of homogeneity. However, in many practical situations the experimenter's goal is to *select* the "best" treatment; a test of homogeneity does not provide the information which the experimenter truly seeks in such situations. This paper provides an appropriate formulation of the selection problem and gives a procedure for attaining this goal.

2. ASSUMPTIONS, NOTATION AND PROBLEM FORMULATION

Consider two treatments T_1 and T_2 and let π_{ij} denote the probability that a matched observation on T_1 and T_2 results in outcome i with T_1 and outcome j with T_2 $(i,j = 0, 1)$ where 1 denotes success and 0 denotes failure. We have $\Sigma\Sigma\pi_{ij} = 1$. We assume that the π_{ij} remain constant throughout the trial. Thus each matched observation can be thought of as a realization from a fixed multinomial distribution with four cells: (1, 1), (1, 0), (0, 1) and (0, 0); the corresponding probabilities are π_{11}, π_{10}, π_{01} and π_{00}, respectively. Let $p_1 = \pi_{11}+\pi_{10}$ and $p_2 = \pi_{11}+\pi_{01}$ be the success probabilities of T_1 and T_2 respectively and let $p_{[1]} \leqslant p_{[2]}$ denote the ordered values of the p_i. We assume that the π_{ij} are unknown, but the experimenter is able to *specify* an upper limit $\pi^*(0 < \pi^* \leqslant 1)$ on $\pi_{10}+\pi_{01} = \pi$ (say). (In general, if matching is properly done and T_1 and T_2 are comparable to each other then π, the probability of different outcomes on T_1 and T_2 with the same matched observation, will be small; see Section 5 for further discussion.) The experimenter's *goal* is to select the treatment associated with $p_{[2]}$; such a selection is referred to as a *correct selection* (CS) and the corresponding probability is denoted by PCS. The experimenter restricts consideration to procedures which guarantee the *probability requirement*:

$$\text{PCS} \geqslant P^* \quad \text{whenever } p_{[2]}-p_{[1]} = \delta \geqslant \delta^*, \quad \text{and} \quad \pi_{10}+\pi_{01} = \pi \leqslant \pi^*, \tag{2.1}$$

where $\{\pi^*, \delta^*, P^*\}$ are constants specified before experimentation starts; $0 < \pi^* \leqslant 1$, $0 < \delta^* \leqslant \pi^*$ and $\frac{1}{2} < P^* < 1$.

3. The Proposed Procedure and its PCS

We propose a natural selection procedure which takes n matched observations on T_1 and T_2. Let the outcomes be represented in a 2×2 table:

		T_2	
		Success	Failure
T_1	Success	x_{11}	x_{10}
	Failure	x_{01}	x_{00}

The decision rule is: select T_1 if $x_{10}>x_{01}$; select T_2 if $x_{10}<x_{01}$; and select T_1 or T_2 at random assigning equal probability to each one if $x_{10}=x_{01}$. It might be noted that x_{11} and x_{00} play no role in the selection procedure. Our main problem is to find the minimum value of n which guarantees (2.1).

Without loss of generality assume that T_1 is the better treatment, i.e. $p_1>p_2$ or equivalently $\pi_{10}>\pi_{01}$. Then

$$\begin{aligned}\text{PCS} &= P\{X_{10}>X_{01}\}+\tfrac{1}{2}P\{X_{10}=X_{01}\}\\ &= \sum_{x=0}^{n}[P\{X_{10}>X_{01}\,|\,X_{10}=x\}+\tfrac{1}{2}P\{X_{10}=X_{01}\,|\,X_{10}+X_{01}=x\}]\binom{n}{x}\pi^x(1-\pi)^{n-x}. \quad (3.1)\end{aligned}$$

Note that the quantity inside the square brackets in (3.1) is just $P(Y>\frac{1}{2}x)+\frac{1}{2}P(Y=\frac{1}{2}x)$ if x is even, and $P\{Y\geqslant\frac{1}{2}(x+1)\}$ if x is odd, where Y has a binomial distribution with parameters x and $\lambda=\pi_{10}/\pi$. Denoting it by $g(x,\lambda)$ we have

$$g(x,\lambda)=I_\lambda\{\tfrac{1}{2}(x+1),\tfrac{1}{2}(x+1)\} \quad \text{for odd } x\geqslant 1, \quad (3.2a)$$

$$=I_\lambda(\tfrac{1}{2}x,\tfrac{1}{2}x) \quad \text{for even } x\geqslant 2, \quad (3.2b)$$

$$=\tfrac{1}{2} \quad \text{for } x=0, \quad (3.2c)$$

where

$$I_p(a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\int_0^p u^{a-1}(1-u)^{b-1}\,du$$

denotes the usual incomplete beta function.

To guarantee (2.1) it is necessary to find the infimum of the PCS over the region $p_1-p_2=\pi_{10}-\pi_{01}\geqslant\delta^*$, $\pi_{10}+\pi_{01}\leqslant\pi^*$; the minimum value of n which makes this infimum $\geqslant P^*$ will be the desired sample size. To find the infimum, represent $\text{PCS}=E_\pi\{g(X,\lambda)\}$ where X is a binomial random variable with parameters n and π and E_π denotes the expectation evaluated at parameter value π. Intuitively the infimum of the PCS over the specified region will occur when the p_i are as close as possible and the matching is as ineffective as possible, i.e. when $p_1-p_2=\pi_{10}-\pi_{01}=\delta^*$ and $\pi_{10}+\pi_{01}=\pi^*$ which is referred to as the least favourable configuration (LFC). However, note that it is not completely obvious that the PCS decreases with increasing π since this also corresponds to an increase in the number of "effective" observations x_{10} and x_{01}. A formal proof of the LFC is thus needed and is given in the Appendix. Note that at the LFC, we have

$$\begin{aligned}\text{PCS}_{\text{LFC}} &= E_{\pi^*}\{g(X,\tfrac{1}{2}+\delta^*/2\pi^*)\}\\ &= \sum_{x=0}^{n} g(x,\tfrac{1}{2}+\delta^*/2\pi^*)\binom{n}{x}(\pi^*)^x(1-\pi^*)^{n-x}. \quad (3.3)\end{aligned}$$

It is fairly straightforward to verify that the right-hand side of (3.3) is increasing in n and tends to 1 as n tends to ∞. Thus any desired value of P^* can be attained by choosing n large enough. It should be noted that, when $\delta^*=\pi^*$, (3.3) simplifies to $1-\frac{1}{2}(1-\pi^*)^n$ and when $\pi^*=1$, (3.3) simplifies to $g(n,\frac{1}{2}+\frac{1}{2}\delta^*)$.

4. Tables of Sample Sizes

The values of n which guarantee (2.1) were found using (3.3) for $n \leqslant 35$. For $n > 35$, the following normal approximation was used. Note that since X_{10}, X_{01} are multinomial frequencies, for large n, $(X_{10} - X_{01})$ can be regarded as a normal random variable with mean $= n(\pi_{10} - \pi_{01})$ and variance $= n\{\pi_{10} + \pi_{01} - (\pi_{10} - \pi_{01})^2\}$. Therefore the PCS under the LFC can be written as $\text{PCS}_{\text{LFC}} \cong \Phi\{\delta^* \sqrt{n} / \sqrt{(\pi^* - \delta^{*2})}\}$ where $\Phi(.)$ denotes the standard normal distribution function. From this we obtain

$$n \cong \frac{(\pi^* - \delta^{*2})\{\Phi^{-1}(P^*)\}^2}{\delta^{*2}}. \tag{4.1}$$

This approximation is useful when P^* is large and/or δ^* is small and/or π^* is large. The values of n obtained from (4.1) were rounded upwards. The calculations for $P^* = 0{\cdot}90$ and $0{\cdot}95$ appear in Table 1; the values of n for $P^* = 0{\cdot}99$ are almost exactly double those for $P^* = 0{\cdot}95$ and hence are not given here.

TABLE 1

Values of n

$P^* = 0{\cdot}90$

π^* \ δ^*	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
0·1	65	16								
0·2	130	32	14	8						
0·3	196	48	21	12	7	5				
0·4	262	65	29	16	10	7	5	4		
0·5	327	81	35	20	13	9	6	5	4	3
0·6	393	97	43	24	15	10	8	6	4	3
0·7	459	114	50	28	18	12	9	7	5	4
0·8	524	130	57	32	21	14	10	8	6	5
0·9	590	147	65	36	23	16	12	9	7	5
1·0	656	163	72	40	25	17	13	9	7	7

$P^* = 0{\cdot}95$

π^* \ δ^*	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
0·1	106	23								
0·2	214	52	21	11						
0·3	322	79	34	18	11	7				
0·4	431	106	46	25	16	10	7	5		
0·5	539	133	58	32	20	14	10	7	5	4
0·6	647	160	70	38	25	17	12	9	7	5
0·7	755	187	82	45	29	20	14	11	8	6
0·8	864	214	94	52	34	23	17	13	10	8
0·9	972	241	106	59	37	26	19	14	11	9
1·0	1080	268	118	65	41	29	21	17	13	9

To check the accuracy of the normal approximation we computed exact and approximate n-values for $20 \leqslant n \leqslant 35$ and found that the approximate n is always within ± 1 of the exact n; the accuracy of the approximation improves with increasing n. Thus the normal approximation should be very good for $n > 35$.

5. Comparison of Matched Samples Design with Independent Samples Design

In the case of independent samples design the *goal* of the experimenter is the same as before, namely to select the treatment having the larger success probability. However, since no π_{ij} are present here, the probability *requirement* simply reads

$$\text{PCS} \geqslant P^* \quad \text{whenever } p_{[2]} - p_{[1]} = \delta \geqslant \delta^*, \tag{5.1}$$

where $\{\delta^*, P^*\}$ are constants specified before experimentation starts; $0 < \delta^* < 1$ and $\frac{1}{2} < P^* < 1$. Sobel and Huyett (1957) proposed a single-stage procedure which guarantees (5.1) and showed that the optimal sample size per population, in large samples, is

$$n \cong \frac{(1-\delta^{*2})\{\Phi^{-1}(P^*)\}^2}{2\delta^{*2}}. \tag{5.2}$$

The relative efficiency (RE) of the matched samples design in terms of the ratio of sample sizes obtained from (4.1) and (5.2) is

$$\text{RE} \cong \frac{1-\delta^{*2}}{2(\pi^* - \delta^{*2})}. \tag{5.3}$$

Note that we assume the values of δ^* and P^* specified by (2.1) and (5.1) are the same and that RE does not depend on P^*. Furthermore, $\text{RE} > 1$ if $\pi^* < \frac{1}{2}(1+\delta^{*2})$. Also if $\pi^* = 1$ (i.e. $\pi = \pi_{10} + \pi_{01}$ is not constrained by any prior knowledge about its value) then $\text{RE} = 0{\cdot}5$. Thus for $\pi^* > \frac{1}{2}(1+\delta^{*2})$, the matched samples design is less efficient in the large sample case than the independent samples design. If the experimenter does not assume any prior knowledge concerning the value of π, then in the large sample case the matched samples design requires twice as many observations as the independent samples design to guarantee the same probability requirement.

These results give a quantitative measure of how effective the matching must be (in our notation how small π must be) so that the matched samples design is more efficient than the independent samples design. The main conclusion to be drawn from this discussion is that, in the design of a matched samples experiment a high level of matching should be ensured. This can be achieved by choosing the matching variables so that they are highly correlated with the outcome variables. If the matching is ineffective then there can be considerable loss in efficiency relative to the independent samples design.

The results obtained here are in broad agreement with the similar work done for the testing problem in 2×2 tables by several authors, see, for example, Youkeles (1963), Worcester (1964) and Miettinen (1968). These authors also reach the conclusion that for testing the homogeneity of two proportions, matched samples design can be disadvantageous if the matching is not effective and the advantage is not substantial unless the matching is highly effective. For additional references and also for some practical aspects of matching see McKinlay (1977).

Acknowledgements

The author is thankful to three referees for making many useful suggestions for improvement of the presentation. This research was supported by NSF Grant No. ENG77-06112.

References

Bechhofer, R. E. (1954). A single-sample multiple-decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, **25**, 16–39.

Bennet, B. M. (1967). Tests of hypotheses concerning matched samples. *J. R. Statist. Soc.* B, **29**, 468–474.

—— (1968). Note on χ^2 tests for matched samples. *J. R. Statist. Soc.* B, **30**, 368–370.

Bhapkar, V. P. (1973). On the comparison of proportions in matched samples. *Sankhyā* A, **35**, 341–356.

Bhapkar, V. P. and Somes, G. W. (1977). Distribution of Q when testing equality of matched proportions. *J. Amer. Statist. Ass.*, **72**, 658–661.

COCHRAN, W. G. (1950). Comparison of percentages in matched samples. *Biometrika*, **37**, 256–266.
GUPTA, S. S. and PANCHAPAKESAN, S. (1972). On a class of subset selection procedures. *Ann. Math. Statist.*, **43**, 814–822.
MCKINLAY, S. M. (1977). Pair-matching—a reappraisal of a popular technique. *Biometrics*, **33**, 725–735.
MCNEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–157.
MIETTINEN, O. S. (1968). The matched pairs design in the case of all-or-none response. *Biometrics*, **24**, 339–352.
SOBEL, M. and HUYETT, M. J. (1957). Selecting the best one of several binomial populations. *Bell System Tech. J.*, **36**, 537–576.
WORCESTER, J. (1964). Matched samples in epidemiologic studies. *Biometrics*, **20**, 840–848.
YOUKELES, L. H. (1963). Loss of power through ineffective pairing of observations on small two-treatment all-or-none experiments. *Biometrics*, **19**, 175–180.

APPENDIX

To prove the assertion regarding the LFC, we first keep π fixed and regard the PCS as a function of $\lambda = \pi_{10}/\pi$. To show that, subject to $\pi_{10}+\pi_{01} = \pi$ (fixed), the PCS is minimized at $\pi_{10}-\pi_{01} = \delta^*$, it suffices to show that g is a non-decreasing function of λ for each x. But this follows immediately from (3.2); in fact g is a strictly increasing function of $\lambda > \frac{1}{2}$ for each $x \geqslant 1$.

Our next task is to find the infimum over $\pi \leqslant \pi^*$ of

$$\inf_{\pi_{10}-\pi_{01} \geqslant \delta^*} \text{PCS} = E_\pi\{g(X, \lambda^*)\}, \tag{A.1}$$

where $\lambda^* = \frac{1}{2}(\pi+\delta^*)/\pi$. To show that this infimum occurs at $\pi = \pi^*$ we prove the following.

Theorem. $E_\pi(g(X, \lambda^*)\}$ is a decreasing function of π.

Proof. Denote the binomial probability function $\binom{n}{x} p^x(1-p)^{n-x}$ by $b(x; n,p)$ and the corresponding distribution function by $B(x; n,p)$. A "discrete analog" of Theorem 2.1 of Gupta and Panchapakesan (1972) shows that the condition to be verified for the monotonicity of $E_\pi(g(X, \lambda^*)\}$ relative to π is

$$\{(\partial/\partial\pi)\, g(x, \lambda^*)\}\, b(x; n, \pi) - \{g(x, \lambda^*) - g(x-1, \lambda^*)\}\, (\partial/\partial\pi)\, B(x-1; n, \pi) \leqslant 0 \tag{A.2}$$

for $1 \leqslant x \leqslant n$; for $x = 0$ the left-hand side of (A.2) is 0. Substitute in (A.3),

$$(\partial/\partial\pi)\, B(x-1; n, \pi) = -(x/\pi)\, b(x; n, \pi) \tag{A.3}$$

and find that the condition to be verified becomes

$$(\partial/\partial\pi) g(x, \lambda^*) + (x/\pi)\{g(x, \lambda^*) - g\,(x-1, \lambda^*)\} \leqslant 0. \tag{A.3}$$

With some algebraic manipulations it can be shown that, for x odd,

$$g(x, \lambda^*) - g(x-1, \lambda^*) = -(\pi/x)\, (\partial/\partial\pi)\, g(x, \lambda^*).$$

Thus, we find that the left-hand side of (A.3) is 0 for x odd. For x even $\geqslant 2$ it follows from (3.2a) and (3.2b) that $g(x, \lambda^*) - g(x-1, \lambda^*) = 0$. Furthermore, it can be easily verified that $(\partial/\partial\pi)\, g(x, \lambda^*) < 0$. Thus (A.3) is verified in all the cases. Because of the strict inequality for x even $\geqslant 2$, it follows that $E_\pi\{g(X, \lambda^*)\}$ is strictly decreasing in π.