SMALL SAMPLE RESULTS FOR THE 27% RULE

by

George P. McCabe, Jr., Purdue University

Department of Statistics
Division of Mathematical Sciences
September 1978

SMALL SAMPLE RESULTS FOR THE 27% RULE

by

George P. McCabe, Jr., Purdue University

I. MOTIVATING PROBLEM

In social science research the following type of problem is occasionally encountered. A fairly large collection of individuals, e.g. all students in an introductory psychology class, are measured on a variable, denoted by X. The observations are ordered and two groups are formed: one corresponding to low-X scores and the other corresponding to high-X. Sometimes the split is performed at the median; in other instances, the upper and lower thirds are used. In the latter case, no further observations are taken on the individuals in the middle third of the collection.

The designations low-X and high-X are used subsequently as a dichotomous variable, with the actual X-score being ignored. This dichotomous variable is then treated as a two-level factor in an experimental design. The simplest case of such a design, which will be the only one studied here in detail, is when a single additional variable, denoted by Y, is measured. The two sample t-test is then used to compare the performance of the low-X and high-X groups on Y.

The purpose of this investigation is not to defend the appropriateness of the above procedure. Rather, the properties of the procedure are examined and some useful information for determining efficient low-X and high-X groups is given.

II. MODEL ASSUMPTIONS AND NOTATION

Let (X_i,Y_i), $i = 1,\ldots,N$ be bivariate normal random variables. In what follows, it can be assumed without loss of generality, that the means are zero and the variances are one. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)}$ denote the order statistics of the X variable. For any $n \leq N/2$, let

$$X_{1j} = X_{(j)}, \quad \text{for } j = 1,\ldots,n$$

and

$$X_{2j} = X_{(n-j+1)}, \text{ for } j = 1,\ldots,n$$

The Y observation paired with X_{ij} will be denoted by Y_{ij} for $i = 1,2$ and $j = 1,\ldots,n$.

Let be the fraction of observations in each tail to be designated low-X and high-X. Thus, we let $\alpha \in (0.5]$ and define n by $n = [\alpha N]$, where $[\cdot]$ is the greatest integer function. The sample means and variances for the Y variable are

$$\bar{Y}_i(\alpha) = n^{-1} \sum_{i=1}^{n} Y_{ij}, \; i = 1,2$$

and

$$s_i^2(\alpha) = (n-1)^{-1} \sum_{i=1}^{n} (Y_{ij}-\bar{Y}_i(\alpha))^2, \; i = 1,2.$$

The t-statistic for comparing the Y means of the low-X and high-X groups is

$$t(\alpha) = \frac{\bar{Y}_1(\alpha) - \bar{Y}_2(\alpha)}{\sqrt{\dfrac{s_1^2(\alpha) + s_2^2(\alpha)}{n-1}}}.$$

III. LARGE SAMPLE RESULTS

In [1] it is shown that

$$\lim_{N\to\infty} N^{-1}t^2(\alpha) = \frac{2\rho^2(1-\Phi(a))}{M^2(a)+\rho^2(aM(a)-1)},$$

where ρ is the correlation between X and Y, $\Phi(\cdot)$ is the standard normal cumulative distribution function, a is defined by $\alpha = 1-\Phi(a)$ and $M(\cdot)$ is the Mills ratio, $M(x) = (1-\Phi(x))/\varphi(x)$, with $\varphi(\cdot)$ being the standard normal density function.

For a given value of ρ, the value of α which maximizes the above expression can be found. Details are given in [1]. As $|\rho^2|$ approaches zero, the optimal α approaches 27% from below. For $|\rho^2| = .5$ the optimal value is about 24% and falls off to about 20% for $|\rho^2| = .8$. For $|\rho^2| = .95$, $\alpha = 16\%$.

In summary, the large sample results indicate that a choice of $\alpha = 25\%$ is effective for a reasonable range of values of ρ that one might expect to encounter in practice.

IV. SMALL SAMPLE RESULTS

Of course, choosing α which maximizes the limiting value of $N^{-1}t^2(\alpha)$ is not equivalent to finding the α which maximizes the power of the t-test for detecting nonzero values of ρ. In addition, the relevance of the asymptotic calculations for reasonable size samples must be examined.

To address these questions, several simulations were run. Values of N chosen were 10, 20, 30, 40, 50, 60 and 100. For the first four values of N, 10,000 simulations were run; for the next two, 5,000 and for the last 4,000. The following values of $|\rho^2|$ were used: 0, .05, .1, .2, .3, .4, .5, .6, .7, .8, .9, .99. The five percent and one percent powers were estimated for all possible values of α.

In the simulations, the same generated random variables were used for all values of $|\rho^2|$ by considering the appropriate conditional distributions. The normal random numbers were generated using the routine described in [2].

Inspection of the results of the simulations reveals that the large sample results are applicable for practical values of N, i.e. it is reasonable to choose $\alpha = 25\%$ for cases where $|\rho^2|$ is expected to be small or moderate and slightly lower values of α when $|\rho^2|$ is expected to be large.

A very interesting fact revealed by the simulations is that the power as a function of α is very flat. The difference in performance between the optimal α and close values is often negligible from a practical point of view. This observation led to the construction of Tables 1 and 2. In Table 1 values of α^* are given for all values of $|\rho^2|$ and N considered. The quantity α^* is defined to be the smallest α giving power that is not less than .01 less than the power of the optimal α, where power is the power of the t-test using a type I error of 5%. Table 2 gives the powers for these α^*. Results for a Type I error of 1% are qualitatively similar.

Since observations on the variable Y are taken on $2n = 2[\alpha N]$ cases, substantial savings can result by choosing α as small as possible while still retaining good power. As can be seen from Table 2, it is often possible to have excellent power with very few observations. For example, with $N = 50$ and $|\rho^2| = .5$, one only needs $\alpha = 8\%$, i.e. $n = 4$ to get 99% power (rounded to 2 places.)

REFERENCES

[1] McCabe, George P. Jr. (1977) Use of the 27% rule in experimental design. Purdue University Department of Statistics Mimeo Series No. 499.

[2] Rubin, Herman (1974) RVP-Random variable package. Purdue University Computing Center.

TABLE 1

Values of α^*

	α^*(%)	N 10	20	30	40	50	60	100
	.05	30	25	23.3	22.5	24	23.3	27
	.10	30	25	26.7	25	24	25	22
	.20	30	30	23.3	25	22	21.7	13
	.30	30	30	23.3	22.5	20	15	7
	.40	40	30	23.3	15	12	10	4
$\|\rho^2\|$	.50	40	25	20	12.5	8	6.7	3
	.60	40	25	20	10	8	5	3
	.70	40	20	10	7.5	6	5	3
	.80	40	15	10	7.5	6	3.3	2
	.90	30	15	6.7	5	4	3.3	2
	.99	30	10	6.7	5	4	3.3	2

TABLE 2

Power of the 5% t-test for $\alpha = \alpha^*$

	Power (%)	N 10	20	30	40	50	60	100
	.05	13	20	27	33	40	45	66
	.10	19	32	44	55	64	71	89
	.20	28	54	70	82	90	94	99
	.30	39	72	87	95	98	99	99
	.40	51	85	95	98	99	100	99
$\|\rho^2\|$	.50	62	93	99	99	99	100	99
	.60	73	98	100	100	100	100	100
	.70	83	99	100	100	100	100	100
	.80	92	99	100	100	100	100	100
	.90	98	100	100	100	100	100	100
	.99	100	100	100	100	100	100	100

SMALL SAMPLE RESULTS FOR THE 27% RULE

George P. McCabe, Jr., Purdue University

I. MOTIVATING PROBLEM

In social science research the following type of problem is occasionally encountered. A fairly large collection of individuals, e.g. all students in an introductory psychology class, are measured on a variable, denoted by X. The observations are ordered and two groups are formed: one corresponding to low-X scores and the other corresponding to high-X. Sometimes the split is performed at the median; in other instances, the upper and lower thirds are used. In the latter case, no further observations are taken on the individuals in the middle third of the collection.

The designations low-X and high-X are used subsequently as a dichotomous variable, with the actual X-score being ignored. This dichotomous variable is then treated as a two-level factor in an experimental design. The simplest case of such a design, which will be the only one studied here in detail, is when a single additional variable, denoted by Y, is measured. The two sample t-test is then used to compare the performance of the low-X and high-X groups on Y.

The purpose of this investigation is not to defend the appropriateness of the above procedure. Rather, the properties of the procedure are examined and some useful information for determining efficient low-X and high-X groups is given.

II. MODEL ASSUMPTIONS AND NOTATION

Let (X_i,Y_i), $i = 1,\ldots,N$ be bivariate normal random variables. In what follows, it can be assumed without loss of generality, that the means are zero and the variances are one. Let $X_{(1)} \leq X_{(2)} \leq\ldots\leq X_{(N)}$ denote the order statistics of the X variable. For any $n \leq N/2$, let

$$X_{1j} = X_{(j)}, \quad \text{for } j = 1,\ldots,n$$

and

$$X_{2j} = X_{(n-j+1)}, \text{ for } j = 1,\ldots,n$$

The Y observation paired with X_{ij} will be denoted by Y_{ij} for $i = 1,2$ and $j = 1,\ldots,n$.

Let α be the fraction of observations in each tail to be designated low-X and high-X. Thus, we let $\alpha \in (0.5]$ and define n by $n = [\alpha N]$, where $[\cdot]$ is the greatest integer function. The sample means and variances for the Y variable are

$$\bar{Y}_i(\alpha) = n^{-1} \textstyle\sum_{i=1}^{n} Y_{ij}, \quad i = 1,2$$

and

$$s_i^2(\alpha) = (n-1)^{-1}\textstyle\sum_{i=1}^{n}(Y_{ij}-\bar{Y}_i(\alpha))^2, \; i=1,2.$$

The t-statistic for comparing the Y means of the low-X and high-X groups is

$$t(\alpha) = \frac{\bar{Y}_1(\alpha) - \bar{Y}_2(\alpha)}{\sqrt{\dfrac{s_1^2(\alpha) + s_2^2(\alpha)}{n-1}}}\,.$$

III. LARGE SAMPLE RESULTS

In [1] it is shown that

$$\lim_{N\to\infty} N^{-1}t^2(\alpha) = \frac{2\rho^2(1-\Phi(a))}{M^2(a)+\rho^2(aM(a)-1)},$$

where ρ is the correlation between X and Y, $\Phi(\cdot)$ is the standard normal cumulative distribution function, a is defined by $\alpha = 1-\Phi(a)$ and $M(\cdot)$ is the Mills ratio, $M(x) = (1-\Phi(x))/\varphi(x)$, with $\varphi(\cdot)$ being the standard normal density function.

For a given value of ρ, the value of α which maximizes the above expression can be found. Details are given in [1]. As $|\rho^2|$ approaches zero, the optimal α approaches 27% from below. For $|\rho^2| = .5$ the optimal value is about 24% and falls off to about 20% for $|\rho^2| = .8$. For $|\rho^2| = .95$, $\alpha = 16\%$.

In summary, the large sample results indicate that a choice of $\alpha = 25\%$ is effective for a reasonable range of values of ρ that one might expect to encounter in practice.

IV. SMALL SAMPLE RESULTS

Of course, choosing α which maximizes the limiting value of $N^{-1}t^2(\alpha)$ is not equivalent to finding the α which maximizes the power of the t-test for detecting nonzero values of ρ. In addition, the relevance of the asymptotic calculations for reasonable size samples must be examined.

To address these questions, several simulations were run. Values of N chosen were 10, 20, 30, 40, 50, 60 and 100. For the first four values of N, 10,000 simulations were run; for the next two, 5,000 and for the last 4,000. The following values of $|\rho^2|$ were used: 0, .05, .1, .2, .3, .4, .5, .6, .7, .8, .9, .99. The five percent and one percent powers were estimated for all possible values of α.

In the simulations, the same generated random variables were used for all values of $|\rho^2|$ by considering the appropriate conditional distributions. The normal random numbers were generated using the routine described in [2].

Inspection of the results of the simulations reveals that the large sample results are applicable for practical values of N, i.e. it is reasonable to choose $\alpha = 25\%$ for cases where $|\rho^2|$ is expected to be small or moderate and slightly lower values of α when $|\rho^2|$ is expected to be large.

A very interesting fact revealed by the simulations is that the power as a function of α is very flat. The difference in performance between the optimal α and close values is often negligible from a practical point of view. This observation led to the construction of Tables 1 and 2. In Table 1 values of α^* are given for all values of $|\rho^2|$ and N considered. The quantity α^* is defined to be the smallest α giving power that is not less than .01 less than the power of the optimal α, where power is the power of the t-test using a type I error of 5%. Table 2 gives the powers for these α^*. Results for a Type I error of 1% are qualitatively similar.

Reprinted from the 1978 Social Statistics Section Proceedings of the American Statistical Association.

Since observations on the variable Y are taken on $2n = 2[\alpha N]$ cases, substantial savings can result by choosing α as small as possible while still retaining good power. As can be seen from Table 2, it is often possible to have excellent power with very few observations. For example, with $N = 50$ and $|\rho^2| = .5$, one only needs $\alpha = 8\%$, i.e. $n = 4$ to get 99% power (rounded to 2 places.)

REFERENCES

[1] McCabe, George P. Jr. (1977) Use of the 27% rule in experimental design. Purdue University Department of Statistics Mimeo Series No. 499.

[2] Rubin, Herman (1974) RVP-Random variable package. Purdue University Computing Center.

Table 1

Values of α^*

	α^*(%)	N: 10	20	30	40	50	60	100
	.05	30	25	23.3	22.5	24	23.3	27
	.10	30	25	26.7	25	24	25	22
	.20	30	30	23.3	25	22	21.7	13
	.30	30	30	23.3	22.5	20	15	7
	.40	40	30	23.3	15	12	10	4
$\|\rho^2\|$	.50	40	25	20	12.5	8	6.7	3
	.60	40	25	20	10	8	5	3
	.70	40	20	10	7.5	6	5	3
	.80	40	15	10	7.5	6	3.3	2
	.90	30	15	6.7	5	4	3.3	2
	.99	30	10	6.7	5	4	3.3	2

Table 2

Power of the 5% t-test for $\alpha = \alpha^*$

	Power(%)	N: 10	20	30	40	50	60	100
	.05	13	20	27	33	40	45	66
	.10	19	32	44	55	64	71	89
	.20	28	54	70	82	90	94	99
	.30	39	72	87	95	98	99	99
$\|\rho^2\|$	.40	51	85	95	98	99	100	99
	.50	62	93	99	99	99	100	99
	.60	73	98	100	100	100	100	100
	.70	83	99	100	100	100	100	100
	.80	92	99	100	100	100	100	100
	.90	98	100	100	100	100	100	100
	.99	100	100	100	100	100	100	100