International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(3): 105-109

https://www.mathsjournal.com
Received: 15-03-2024
Accepted: 21-04-2024

Bayan Shukr
Department of Statistics, North Dakota State University
Fargo, North Dakota

Rhonda Magel
Department of Statistics, North Dakota State University
Fargo, North Dakota

Corresponding Author:
Bayan Shukr
Department of Statistics, North Dakota State University
Fargo, North Dakota

Tests of two proportions for use in a mixed design

Bayan Shukr and Rhonda Magel

Abstract
Five tests for two proportions are proposed for use with data in a mixed design consisting of paired and independent data. The tests are combinations of the standardized versions of the McNemar test and the two-sample independent test for proportions using different weights. A simulation study is conducted comparing the estimated powers of the tests when the sample size for the paired data is equal to the sample sizes of the independent sample data (assuming equal independent sample sizes); when the sample size for the paired data is smaller than the sample sizes for the independent sample data; and when the sample sizes for the independent sample data are smaller than the sample size for the paired data. Correlations of 0.2, 0.3, and 0.4 are considered in the paired data. It is found that the test having higher powers when the paired sample size is larger is the test that assigns equal weights to both standardized versions or the test that assigns twice as much weight to the two-sample independent test. When the independent sample sizes are larger than the paired sample size, it is found the test having higher powers is the one that assigns more weight to the McNemar test. When the sample sizes are equal, the test that has higher powers assigns equal weights to both.

Keywords: McNemar test, two-sample test for proportions, paired data, independent data

Introduction

In this research, we will focus on testing for differences between proportions, specifically whether the treatment group has a significantly higher proportion of successes than the control group. We are proposing tests in this case for a mixed design consisting of both paired and independent data. This type of design may occur if a researcher started with a paired design, ended up with not enough pairs, and then used an independent two-sample design. This is a similar type of problem considered by Dubnicka, Blair, and Hettmansperger (2002) [1] in testing for means with paired and independent sample data.

Our interest lies in determining if there is evidence that the treatment is more effective than the control method and test the null hypothesis and alternative hypothesis below:

$$H_0: p_1 = p_2 \tag{1}$$

$$H_\alpha: p_1 < p_2$$

In this case, p_1 is the proportion of successes with the control and p_2 is the proportion of successes with the new treatment, and we assume we have both paired and independent sample data.

As an example of when we may want to use this test, a treatment for a type of cancer is currently used. Another treatment is being introduced. Researchers want to see if the five-year survival rate for the new treatment is better than what is currently being used. In a clinical trial, patients with this type of cancer are paired up by age, stage of cancer, and other similarities, and one patient of each pair is randomly assigned to the old treatment and one to the new treatment. There may not be enough similar pairs, so several people are left, and these people are randomly assigned to one of the two treatments. Researchers want to test whether the five-year survival rate for the new treatment is significantly higher than for the old treatment.

Finite element formulation oxides necessary flexibility in taking care of different behavior of distinctly different subregions.

Five tests are proposed to test proportions in a mixed design. A simulation study is conducted comparing the five tests on the basis of estimated powers. All five tests use weighted versions of McNemar's test for paired data and the two independent sample tests for proportions.

McNemar's Test

The McNemar's test is a statistical test used to test for differences in proportions when data are paired and binary (McNemar, 1947) [2]. One person in each pair is randomly assigned the treatment and the other person, the control. Binary data is collected from each person indicating the treatment or control was a success or failure. The data are typically coded as 0's and 1's.

The test statistic for McNemar's is given by:

$$McNemar^* = \frac{B-C}{\sqrt{B+C}}$$

Where

B is the number of cases in which the member of the pair receiving the control had a success and the member of the pair receiving treatment had a failure.

C is the number of cases in which the member of the pair receiving the control had a failure and the member of the pair receiving treatment had a success.

The null hypothesis is rejected when $McNemar^* \leq -Z_\alpha$ at the α level of significance where

Z_α is the (1- α) 100% percentile of the standard normal distribution.

The proportions two-sample Test

The test statistic for testing for two proportions:

$$Z = \frac{(\widehat{p_1}-\widehat{p_2})}{SE}$$

$$SE = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}$$, The standard error

$$\hat{p} = \frac{(n_1\widehat{p_1}+n_2\widehat{p_2})}{n_1+n_2}$$

where: $\widehat{p_1} = \frac{x_1}{n_1}$ and $\widehat{p_2} = \frac{x_2}{n_2}$. and x_1 and x_2 are the number of successes in the two samples, n_1 and n_2 are the sample sizes.

The null hypothesis is rejected when $Z \leq -Z_\alpha$ at the α level of significance where Z_α is the (1- α) 100% percentile of the standard normal distribution.

Materials and Methods

We assumed a mixed design, a type of research design that uses both independent samples data and paired data in the same study. We proposed five tests that combined McNemar's Test and the two-sample independent test in a mixed design. These tests are given below.

1. T_1 gives equal weights to the standardized versions of the independent two-sample test and the McNemar's Test
2. T_2 gives twice as much weight to the standardized version of the independent two-sample test as to the standardized version of McNemar's Test.
3. T_3 gives twice as much weight to the standardized version of McNemar's Test as to the standardized version of independent two-sample test.
4. T_4 multiplies the standardized version of McNemar's Test with corresponding paired sample size and the standardized version of independent two-sample test with corresponding independent sample size (assuming equal independent sample sizes).
5. T_5 multiplies the standardized version of McNemar's Test with corresponding independent samples size (assuming equal sample sizes) and the standardized version of independent two-sample test with corresponding paired sample size.

First Test

$$T_1 = \frac{McNemar^* + Z}{\sqrt{2}} \sim N(0, 1)$$

asymptotic normal distribution with mean 0 and variance 1

Second Test

$$T_2 = \frac{McNemar^* + 2 * Z}{\sqrt{5}} \sim N(0, 1)$$

asymptotic normal distribution with mean 0 and variance 1

Third Test

$$T_3 = \frac{2 * McNemar^* + Z}{\sqrt{5}} \sim N(0, 1)$$

asymptotic normal distribution with mean 0 and variance 1

Fourth Test

$$T_4 = \frac{n_1 * McNemar^* + n_2 * Z}{\sqrt{(n_1)^2+(n_2)^2}} \sim N(0, 1)$$

asymptotic normal distribution with mean 0 and variance 1

n_1 is the number of pairs for the McNemar's Test.

n_2 is the sample size for each of the independent samples and assuming equal sample size.

$N = n_1 + n_2$

Fifth Test

$$T_5 = \frac{n_2 * McNemar^* + n_1 * Z}{\sqrt{(n_1)^2+(n_2)^2}} \sim N(0, 1)$$

asymptotic normal distribution with mean 0 and variance 1

n_1 is the number of pairs for the McNemar's Test.

n_2 is the sample size for each of the independent samples and assuming equal sample size.

$N = n_1 + n_2$

In all of the five tests, H_0 is rejected for a small value which is $T \leq -Z_\alpha$ at the α level of significance where Z_α is the (1- α) 100% percentile of the standard normal distribution. If the test is performed at a 5% level of significance, then Z_α= 1.645

It is noted that if the paired and independent sample sizes are equal, the First Test, is the same as the Fourth and Fifth Test. If the paired sample size is twice as much as each independent sample size, the Third Test and the Fourth Test are the same. If the sample size for the two independent sample test is twice as much as the paired sample size, the Second Test and the Fifth Test are the same.

Example

Consider a business study that seeks to test the success rates of two different marketing strategies: In-Person Sales and Customer Experience Strategy (Strategy A); and Digital Marketing and E-Commerce Strategy (Strategy B). The study uses both independent sample data and paired data allowing researchers to examine whether there is a statistically significant difference in success rates between the two strategies, with the hypothesis that Strategy (B) will be more successful for selling the product. In this case the strategies are:

Strategy A: In-Person Sales and Customer Experience Strategy; and

Strategy B: Digital Marketing and E-Commerce Strategy.

Independent data: A random sample of potential customers is selected and Strategy A is applied. A second random sample of potential customers is selected and Strategy B is applied. In this example n_2= 15 for each of the independent samples

Paired data: Potential customers are paired based on having similar demographic variables (Gender, age, and income). Strategy A is randomly applied to one member of pair and Strategy B is then applied to the other member of each pair where n_1 is the number of pairs. In this example n_1 =60

The binary outcome for each potential customer is whether or not the Strategy applied resulted in a sale (success) (coded as 1) or not (coded as 0).

$H_0: p_A = p_B$
$H_\alpha: p_A < p_B$

Table 1: The independent sample data collected is given below.

	0	1	Total
Strategy (A)	8=0.53	7=0.47	15
Strategy (B)	5=0.33	10=0.67	15

$$Z = \frac{(\underline{p_1} - \underline{p_2})}{SE}$$

$$SE = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}$$, The standard error

$$\hat{p} = \frac{(n_1\underline{p_1} + n_2\underline{p_1})}{n_1 + n_2}$$

Where

$\underline{p_1}(\textit{proportion of success for Strategy } (A)) = \frac{x_1}{n_1} = \frac{7}{15} =$ 0.47 and $\underline{p_2}(\textit{proportion of success for Strategy } (B)) = \frac{x_2}{n_2} = \frac{10}{15} = 0.67$

$$\hat{p} = \frac{(15*0.47+15*0.67)}{30} = 0.57$$

$$SE = \sqrt{0.57(1-0.57)(\frac{1}{15}+\frac{1}{15})} = \sqrt{0.0327} = 0.1808$$

$$Z = \frac{(0.47-0.67)}{0.1808} = \frac{-0.20}{0.1808} = -1.106$$

Table 2: The paired sample data is summarized below.

Strategy (A)	Strategy (B)		Number of samples
0	0	(Failed, failed)	#10
1	0	(Success, failed)	#8
0	1	(Failed, success)	#15
1	1	(Success, success)	#27

Table 3: The paired data was rearranged in the table below so that one could determine the values of B and C as defined earlier.

		Strategy (B)		
		0	1	Total
Strategy (A)	0	A=10	C=15	25
	1	B=8	D=27	35
	Total	18	42	60

$$McNemar^* = \frac{B-C}{\sqrt{B+C}} = \frac{8-15}{\sqrt{8+15}} = -1.46$$

The values of each of the proposed test statistics were calculated as follows along with their corresponding p-values.

Proposed tests

$T_1 = \frac{(McNemar^*+Z)}{\sqrt{2}} = \frac{(-1.106+-1.46)}{\sqrt{2}}$=-1.81 and p-value ≈ .0351

The null hypothesis is rejected when $T_1 \leq -1.645$ in this case, T_1 =-1.81, we will reject H_0, there is significant evidence that Strategy (A)will be less successful than Strategy (B)

$T_2 = \frac{(McNemar^*+2*Z)}{\sqrt{5}} = \frac{(-1.46+2*(-1.106))}{\sqrt{5}}$=-1.64 and p-value ≈ 0.0505

The null hypothesis is rejected when $T_2 \leq -1.645$ in this case, T_2 =-2.694, we do not reject H_0 , there is not significant evidence that Strategy (A) will be less successful than Strategy (B)

$T_3 = \frac{(2*McNemar^*+Z)}{\sqrt{5}} = \frac{(2*(-1.46)+-1.106)}{\sqrt{5}}$=-1.80 and p-value ≈ 0.0359

The null hypothesis is rejected when $T_3 \leq -1.645$ in this case, T_3 =-2.847, we will reject H_0, there is significant evidence that Strategy (A) will be less successful than Strategy (B)

$T_4 = \frac{(15*(McNemar^*)+60*(Z))}{\sqrt{3825}} = \frac{(15*(-1.46)+60*(-1.106))}{\sqrt{3825}}$=-1.40and p-value ≈ 0.0808

The null hypothesis is rejected when $T_4 \leq -1.645$ in this case, T_4 =-2.380, we will reject H_0, there is significant evidence that Strategy (A)will be less successful than Strategy (B)

$T_5 = \frac{(60*(McNemar^*)+15*(Z))}{\sqrt{3825}} = \frac{(60*(-1.46)+15*(-1.106))}{\sqrt{3825}}$=-1.68 and p-value ≈ 0.0465

The null hypothesis is rejected when $T_5 \leq -1.645$ in this case, T_5 =-1.68, we will reject H_0, there is significant evidence that Strategy (A) will be less successful than Strategy (B)

We can conclude in this example that T_1 has lowest p-value with the p-value of T_3 very close. The fifth test also gives significant results, and the second test gave results that were close to being significant. The fourth test had the highest p-value.

Description of Simulation Study

We want to compare the five proposed tests as the basis of estimated powers through a simulation study under a variety of different conditions. Different sample sizes and different correlations will be considered.

To create random independent samples from a Bernoulli distribution, the capacities RAND are utilized in SAS. This requires the user to express the beginning stage "seed". This should be possible utilizing the Call streaminit work prior to utilizing the RAND work. The language structure for this capacity is.

Call streaminit (seed)

In this simulation, we set the seed at 0 to utilize the framework clock. Accordingly, every run of the code generates an alternate set of data. (Bailer, 2010). The Bernoulli distribution's call work is.

RAND ('bernoulli', p)

where *p* is the proportion.

To create random dependent samples from a Bernoulli distribution, paired data are correlated. In the current study, different correlation values were considered using the correlation values of .2, .3, and .4. The study then used SAS to generate (pairing data) multivariate binary variates with a given set of expected values and a specified correlation structure using the modules defined in the preceding sections. We then used the FROOT function to generate the pairing data.

In the FROOT function, the Randmvbinary function implements the Emrich-Piedmonte algorithm by returning an N*d matrix containing zeros and ones when given a probability vector, p, Delta, and a desired correlation matrix. For this study, the desired correlation matrix applied the correlations .2, .3, and .4. Furthermore, each column of the returned matrix is a binary variate, and the simulated data's sample mean, and correlation should be close to the specified parameters (Wicklin, 2013) [4].

All simulations employ 10,000 sample replications. Situations are considered in which the sample size for the number of pairs is equal to the sample size for both independent samples. We only considered when the sample size for both independent samples was the same in all situations considered. The sample sizes used for both paired and independent sample data were 15 and then 20. We next considered simulations in which the paired sample size is higher than the sample size for both independent samples. In this case we considered sample sizes of 30 and 15, 45 and 15, 60 and 15, 70 and 15, and 90 and 15. Simulations are also considered in which the independent sample size for both independent samples are higher than the paired sample size. In this case, we considered sample sizes of 15 and 30, 15 and 45, 15 and 60, 15 and 70, and 15 and 90.

We first estimated alpha in all cases. Alpha was estimated when the null hypothesis was true by simulating 10,000 sets of data, calculating each test statistic, counting the number of times each test statistic rejected divided by 10,000. The stated value of alpha was 0.05. The estimated value of alpha appears on the first line of every table.

We next estimated the powers of all the test statistics by simulating 10,000 sets of data under the conditions given, calculating each test statistic, and counting the number of times each test statistic rejected divided by 10,000.

Results and Discussion

Tables 1 and 2 give results for equal sample sizes of size 15 and correlations of 0.2 and 0.4 for the paired data. Results for equal sample sizes of 20 were similar.

Tables 1 and 2 only give the results for the first three test statistics because the last two will be the same as the first test statistic when sample sizes are equal.

$$T_1 = \frac{McNemar^* + Z}{\sqrt{2}}$$

$$T_2 = \frac{McNemar^* + 2 * Z}{\sqrt{5}}$$

$$T_3 = \frac{2 * McNemar^* + Z}{\sqrt{5}}$$

n_1 is the number of pairs for the McNemar's Test.
n_2 is the sample size for each of the independent samples.

Table 1: Percentage of Rejection for k=2 Populations; when the paired sample size is equal to the independent sample size and the correlation for paired sample is 0.2

n_1	n_2	P_1	P_2	Correlation				Z	$McNemar^*$	T_1	T_2	T_3
15	15	0.5	0.5	1	0.2	0.2	1	0.0537	0.0472	0.0517	0.0557	0.0515
15	15	0.2	0.5	1	0.2	0.2	1	0.5536	0.5914	0.8374	0.7839	0.8158
15	15	0.3	0.5	1	0.2	0.2	1	0.3097	0.3186	0.5169	0.4753	0.4940
15	15	0.4	0.5	1	0.2	0.2	1	0.1393	0.1466	0.2094	0.1991	0.2079

Table 2: Percentage of Rejection for k=2 Populations; when the paired sample size is equal to the independent sample size and the correlation for paired sample is 0.4

n_1	n_2	P_1	P_2	Correlation				Z	$McNemar^*$	T_1	T_2	T_3
15	15	0.5	0.5	1	0.4	0.4	1	0.0510	0.0573	0.0577	0.0567	0.0541
15	15	0.2	0.5	1	0.4	0.4	1	0.5612	0.7449	0.8917	0.8201	0.8981
15	15	0.3	0.5	1	0.4	0.4	1	0.3054	0.4063	0.5713	0.4977	0.5678
15	15	0.4	0.5	1	0.4	0.4	1	0.1385	0.1742	0.2334	0.2088	0.2273

From the results given in Tables 1 and 2, one can see that T_1 generally has the highest estimated powers with the estimated powers of T_3 being close. T_1 weights the standardized versions of the McNemar and the two independent sample test equally. T_3 gives twice as much weight to the standardized McNemar. Results for equal sample sizes of 20 were similar. The correlation in the paired data did not appear to have much of an effect on which test statistic had higher powers.

Table 3 gives the results of the simulation study for when the sample size for the number of pairs is higher than the sample size for both the independent samples. In this case the number of pairs is 4 times higher. When the paired sample size was twice as high as the independent sample size T_1 generally had a little higher powers, but the powers of T_2 were close. The correlation was 0.3 for the paired data. Results were similar for correlations of 0.2 and 0.4 as to which test had higher powers. When the paired sample size was twice as high as the independent sample size, the estimated powers of T_1 were higher than the other test statistics with the estimated powers of T_2 being close. T_1 weights both the McNemar and the two independent sample test equally and T_2 gives more weight to the two independent sample test. When the paired sample size was four times as high as the sample size for the independent sample sizes, T_2 had slightly larger estimated powers than T_1. The same thing was true when the paired sample size was five and six times higher.

Table 4 give the results for when the number of pairs is less than the sample size for each of the independent samples. In Table 4, the paired sample size is 15 and the independent sample size is 60.

An additional observation across both Tests is the minimal impact of correlation on the test results. The changes in correlation values, ranging from 0.2 to 0.4, do not significantly alter the performance of the tests, as far as which test has the better powers under a given situation. This aspect of the study highlights the stability of the tests under different correlation conditions, making the findings relevant for a wide range of practical applications where such variations are common.

Table 3: Percentage of Rejection for k=2 Populations; when the paired sample size is greater than the independent sample size and the correlation for paired sample data is 0.3

n_1	n_2	P_1	P_2	Correlation				Z	$McNemar^*$	T_1	T_2	T_3	T_4	T_5
60	15	0.5	0.5	1	0.3	0.3	1	0.0624	0.0485	0.0514	0.0534	0.0513	0.0492	0.0551
60	15	0.2	0.25	1	0.3	0.3	1	0.1676	0.1148	0.1962	0.1994	0.1744	0.1488	0.1931
60	15	0.2	0.5	1	0.3	0.3	1	0.9712	0.6509	0.9924	0.9937	0.9699	0.9066	0.9890
60	15	0.3	0.5	1	0.3	0.3	1	0.7432	0.3493	0.8184	0.8332	0.7221	0.5815	0.8006
60	15	0.4	0.5	1	0.3	0.3	1	0.3337	0.1470	0.3509	0.3645	0.2906	0.2339	0.3466

Table 4: Percentage of Rejection for k=2 Populations; when the independent sample size is greater than the paired sample size and the correlation for paired sample data is 0.3

n_1	n_2	P_1	P_2	Correlation				Z	$McNemar^*$	T_1	T_2	T_3	T_4	T_5
15	60	0.5	0.5	1	0.3	0.3	1	0.0499	0.0483	0.0510	0.0540	0.0498	0.0538	0.0505
15	60	0.2	0.25	1	0.3	0.3	1	0.1067	0.1917	0.1970	0.1650	0.2144	0.1350	0.2129
15	60	0.2	0.5	1	0.3	0.3	1	0.5541	0.9955	0.9953	0.9540	0.9990	0.8455	0.9989
15	60	0.3	0.5	1	0.3	0.3	1	0.3046	0.8481	0.8550	0.7046	0.9015	0.5362	0.8989
15	60	0.4	0.5	1	0.3	0.3	1	0.1323	0.3671	0.3677	0.2806	0.4110	0.2125	0.4107

Conclusion

Five tests were proposed for the mixed design when testing proportions. It was found that when the sample size for the paired data is equal to the sample size for the independent sample test, it is better to use the test that weights both the standardized versions of the McNemar and the independent sample test equally. When the number of pairs is at least two times larger than the sample size for both of the independent samples, it is better to use the test with more weight on the independent samples test. In this case, T_2 is recommended or T_5. When the independent sample size for both samples is at least two times larger than for the paired sample size, it is better to use the test with more weight on the McNemar test, T_3, or T_5. This study did not find that T_5 had higher powers over the second or third test statistics.

References

1. Dubnicka SR, Blair RC, Hettmansperger TP. Rank-based procedures for mixed paired and two-sample designs. Journal of Modern Applied Statistical Methods. 2002;1(1):32-41.
2. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947;12(2):153-157. doi:10.1007/BF02295996
3. Shafer DS, Zhang Z. Introductory statistics; c2010. p. 465-469.
4. Wicklin R. Simulating Data with SAS. Cary, NC: SAS Institute; c2013. ISBN: 9781629592534.