

McNemar's test

From Wikipedia, the free encyclopedia

In statistics, **McNemar's test** is a statistical test used on paired nominal data. It is applied to 2×2 contingency tables with a dichotomous trait, with matched pairs of subjects, to determine whether the row and column marginal frequencies are equal (that is, whether there is "marginal homogeneity"). It is named after Quinn McNemar, who introduced it in 1947.^[1] An application of the test in genetics is the transmission disequilibrium test for detecting linkage disequilibrium.^[2]

Contents

- 1 Definition
 - 1.1 Variations
- 2 Examples
- 3 Discussion
 - 3.1 Information in the pairings
- 4 Related tests
- 5 See also
- 6 References
- 7 External links

Definition

The test is applied to a 2×2 contingency table, which tabulates the outcomes of two tests on a sample of n subjects, as follows.

	Test 2 positive	Test 2 negative	Row total
Test 1 positive	<i>a</i>	<i>b</i>	<i>a</i> + <i>b</i>
Test 1 negative	<i>c</i>	<i>d</i>	<i>c</i> + <i>d</i>
Column total	<i>a</i> + <i>c</i>	<i>b</i> + <i>d</i>	<i>n</i>

The null hypothesis of marginal homogeneity states that the two marginal probabilities for each outcome are the same, i.e. $p_a + p_b = p_a + p_c$ and $p_c + p_d = p_b + p_d$.

Thus the null and alternative hypotheses are^[1]

$$H_0 : p_b = p_c$$

$$H_1 : p_b \neq p_c$$

Here p_a , etc., denote the theoretical probability of occurrences in cells with the corresponding label.

The McNemar test statistic is:

$$\chi^2 = \frac{(b - c)^2}{b + c}.$$

Under the null hypothesis, with a sufficiently large number of discordants (cells b and c), χ^2 has a chi-squared distribution with 1 degree of freedom. If the χ^2 result is significant, this provides sufficient evidence to reject the null hypothesis, in favour of the alternative hypothesis that $p_b \neq p_c$, which would mean that the marginal proportions are significantly different from each other.

Variations

If either b or c is small ($b + c < 25$) then χ^2 is not well-approximated by the chi-squared distribution. An exact binomial test can then be used, where b is compared to a binomial distribution with size parameter $n = b + c$ and $p = 0.5$. Effectively, the exact binomial test evaluates the imbalance in the discordants b and c . To achieve a two-sided P-value, the P-value of the extreme tail should be multiplied by 2:

$$\text{exact-P-value} = 2 \sum_{i=0}^b \binom{n}{i} 0.5^i (1 - 0.5)^{n-i}$$

which is simply twice the binomial distribution cumulative distribution function with $p = 0.5$ and $n = b + c$.

Edwards^[3] proposed the following continuity corrected version of the McNemar test to approximate the binomial exact-P-value:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}.$$

The mid-P McNemar test (mid-p binomial test) is calculated by subtracting half the probability of the observed b from the exact one-sided P-value, then double it to obtain the two-sided mid-P-value:^{[4][5]}

$$\text{mid-P-value} = 2 \left(\sum_{i=0}^b \binom{n}{i} 0.5^i (1 - 0.5)^{n-i} - 0.5 \binom{n}{b} 0.5^b (1 - 0.5)^{n-b} \right)$$

This is equivalent to:

$$\text{mid-P-value} = \text{exact-P-value} - \binom{n}{b} 0.5^b (1 - 0.5)^{n-b}$$

where the second term is the binomial distribution probability mass function and $n = b + c$. Fortunately, binomial distribution functions are readily available in common software packages and the McNemar mid-P test can easily be calculated.^[5]

The traditional advice has been to use the exact binomial test when $b + c < 25$. However, simulations have shown both the exact binomial test and the McNemar test with continuity correction to be overly conservative.^[5] When $b + c < 6$, the exact-P-value always exceeds the common significance level 0.05. The original McNemar test was most powerful, but often slightly liberal. The mid-P version was almost as powerful as the asymptotic McNemar test and was not found to exceed the nominal significance level.

Examples

In the first example, a researcher attempts to determine if a drug has an effect on a particular disease. Counts of individuals are given in the table, with the diagnosis (disease: *present* or *absent*) before treatment given in the rows, and the diagnosis after treatment in the columns. The test requires the same subjects to be included in the before-and-after measurements (matched pairs).

	After: present	After: absent	Row total
Before: present	101	121	222
Before: absent	59	33	92
Column total	160	154	314

In this example, the null hypothesis of "marginal homogeneity" would mean there was no effect of the treatment. From the above data, the McNemar test statistic:

$$\chi^2 = \frac{(121 - 59)^2}{121 + 59}$$

has the value 21.35, which is extremely unlikely to form the distribution implied by the null hypothesis ($P < 0.001$). Thus the test provides strong evidence to reject the null hypothesis of no treatment effect.

A second example illustrates differences between the asymptotic McNemar test and alternatives.^[5] The data table is formatted as before, with different numbers in the cells:

	After: present	After: absent	Row total
Before: present	59	6	65
Before: absent	16	80	96
Column total	75	86	161

With these data, the sample size (161 patients) is not small, however results from the McNemar test and other versions are different. The exact binomial test gives $P = 0.053$ and McNemar's test with continuity correction gives $\chi^2 = 3.68$ and $P = 0.055$. The asymptotic McNemar's test gives $\chi^2 = 4.55$ and $P = 0.033$ and the mid-P McNemar's test gives $P = 0.035$. Both the McNemar's test and mid-P version provide stronger evidence for a statistically significant treatment effect in this second example.

Discussion

An interesting observation when interpreting McNemar's test is that the elements of the main diagonal do not contribute to the decision about whether (in the above example) pre- or post-treatment condition is more favourable. Thus, the sum $b + c$ can be small and statistical power of the tests described above can be low even though the number of pairs $a + b + c + d$ is large (see second example above).

An extension of McNemar's test exists in situations where independence does not necessarily hold between the pairs; instead, there are clusters of paired data where the pairs in a cluster may not be independent, but independence holds between different clusters.^[6] An example is analyzing the effectiveness of a dental procedure; in this case, a pair corresponds to the treatment of an individual tooth in patients who might have multiple teeth treated; the effectiveness of treatment of two teeth in the same patient is not likely to be independent, but the treatment of two teeth in different patients is more likely to be independent.^[7]

Information in the pairings

John Rice wrote:^[8]

85 Hodgkin's patients [...] had a sibling of the same sex who was free of the disease and whose age was within 5 years of the patient's. These investigators presented the following table:

	Tonsillectomy	No tonsillectomy
Hodgkins	41	44
Control	33	52

They calculated a chi-squared statistic of 1.53, which is not significant.[...] [they] had made an error in their analysis by ignoring the pairings.[...] [their] samples were not independent, because the siblings were paired [...] we set up a table that exhibits the pairings:

		Sibling	
		No tonsillectomy	Tonsillectomy
Patient	No tonsillectomy	37	7
	Tonsillectomy	15	26

It is to the second table that McNemar's test can be applied. Notice that the sum of the numbers in the second table is 85—the number of *pairs* of siblings—whereas the sum of the numbers in the first table is twice as big, 170—the number of individuals. The second table gives more information than the first. The numbers in the first table can be found by using the numbers in the second table, but not vice versa. The numbers in the first table give only the marginal totals of the numbers in the second table.

Related tests

- The binomial sign test gives an exact test for the McNemar's test.
- The Cochran's Q test for two "treatments" is equivalent to the McNemar's test.
- The Liddell's exact test is an exact alternative to McNemar's test.^{[9][10]}
- The Stuart–Maxwell test is different generalization of the McNemar test, used for testing marginal homogeneity in a square table with more than two rows/columns.^{[11][12][13]}
- The Bhapkar's test (1966) is a more powerful alternative to the Stuart–Maxwell test,^{[14][15]} but it tends to be liberal. Competitive alternatives to the extant methods are available.^[16]

See also

- Pearson's chi-squared test

References

1. McNemar, Quinn (June 18, 1947). "Note on the sampling error of the difference between correlated proportions or percentages". *Psychometrika* **12** (2): 153–157. doi:10.1007/BF02295996. PMID 20254758.
2. Spielman RS; McGinnis RE; Ewens WJ (Mar 1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)". *Am J Hum Genet.* **52** (3): 506–16. PMC 1682161. PMID 8447318.
3. Edwards, A (1948). "Note on the "correction for continuity" in testing the significance of the difference between correlated proportions". *Psychometrika* **13**: 185–187. doi:10.1007/bf02289261.
4. Lancaster, H.O. (1961). "Significance tests in discrete distributions.". *J AmStat Assoc* **56**: 223–234. doi:10.1080/01621459.1961.10482105.
5. Fagerland, M.W.; Lydersen, S.; Laake, P. (2013). "The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional". *BMC Medical Research Methodology* **13**: 91. doi:10.1186/1471-2288-13-91.
6. Yang, Z.; Sun, X.; Hardin, J.W. (2010). "A note on the tests for clustered matched-pair binary data". *Biometrical Journal* **52** (5): 638–652. doi:10.1002/bimj.201000035. PMID 20976694.
7. Durkalski, V.L.; Palesch, Y.Y.; Lipsitz, S.R.; Rust, P.F. (2003). "Analysis of clustered matched-pair data". *Statistics in medicine* **22** (15): 2417–28. doi:10.1002/sim.1438. PMID 12872299. Retrieved April 1, 2009.
8. Rice, John (1995). *Mathematical Statistics and Data Analysis* (Second ed.). Belmont, California: Duxbury Press. pp. 492–494. ISBN 0-534-20934-3.
9. Liddell, D. (1976). "Practical Tests of 2×2 Contingency Tables". *Journal of the Royal Statistical Society* **25** (4): 295–304. JSTOR 2988087.
10. "Maxwell's test, McNemar's test, Kappa test". Rimarcik.com. Retrieved 2012-11-22.
11. Sun, Xuezheng; Yang, Zhao (2008). "Generalized McNemar's Test for Homogeneity of the Marginal Distributions" (PDF). SAS Global Forum.
12. Stuart, Alan (1955). "A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification". Biometrika Trust.
13. Maxwell, A.E. (1970). "Comparing the Classification of Subjects by Two Independent Judges" (PDF). The British Journal of Psychiatry.
14. "McNemar Tests of Marginal Homogeneity". John-uebersax.com. 2006-08-30. Retrieved 2012-11-22.
15. Bhapkar, V.P. (1966). "A Note on the Equivalence of Two Test Criteria for Hypotheses in Categorical Data". American Statistical Association.
16. Yang, Z.; Sun, X.; Hardin, J.W. (2012). "Testing Marginal Homogeneity in Matched-Pair Polytomous Data". *Therapeutic Innovation & Regulatory Science* **46** (4): 434–438. doi:10.1177/0092861512442021.

External links

- Vassar College's McNemar 2×2 Grid (<http://faculty.vassar.edu/lowry/propcorr.html>)
- McNemar Tests of Marginal Homogeneity (<http://john-uebersax.com/stat/mcnemar.htm>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=McNemar%27s_test&oldid=685831474"

Categories: Statistical tests | Categorical data | Nonparametric statistics
 | Summary statistics for contingency tables

- This page was last modified on 15 October 2015, at 07:19.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.