BIOS:4120 – Introduction to Biostatistics
Unit 12: Contingency Tables

Knute D. Carter

Department of Biostatistics
The University of Iowa

November 4, 2025

Overview

- The Chi-Square Test
  - 2 x 2 Tables
  - $r \times c$ Tables
- McNemar's Test
- Odds Ratio

2×2 Tables

- Consider drawing samples from two populations (A and B), and measuring a binary ('success'/ 'failure') variable on each of the subjects in the two samples.

- The resulting data could be represented in the form of a table:

	Variable	
Population	Success	Failure
A	n_{11}	n_{12}
B	n_{21}	n_{22}

- Such a table is called a *2×2 Contingency Table*.

- The counts n_{ij} are called *Observed Cell Counts*.

2×2 Tables

- Consider adding row and column totals to the 2×2 table:

	Variable		
Population	Success	Failure	Total
A	n_{11}	n_{12}	n_{1+}
B	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n_{++}

- The row totals n_{1+} and n_{2+}, and the column totals n_{+1} and n_{+2} are called *Marginal Totals*.

2×2 Tables

For such a 2×2 table, consider the following four conditional and two unconditional probabilities:

- $P(S|A) =$ the probability of a success given that a subject is from population A,
- $P(S|B) =$ the probability of a success given that a subject is from population B,
- $P(F|A) =$ the probability of a failure given that a subject is from population A,
- $P(F|B) =$ the probability of a failure given that a subject is from population B,
- $P(S) =$ the probability of a success,
- $P(F) =$ the probability of a failure.

2×2 Tables

	Variable	
Population	Success	Failure
A	$P(S\|A)$	$P(F\|A)$
B	$P(S\|B)$	$P(F\|B)$
	$P(S)$	$P(F)$

- In the context of 2×2 contingency tables, we are often interested in testing the hypothesis that the probability of a success is the same for the first and the second populations.
- Note that this is equivalent to testing that the probability of a failure is the same for the first and the second populations.
- Such a test is called a *Test of Homogeneity*.

2×2 Tables

- Symbolically, we may express the null hypothesis of homogeneous populations as follows:

$$H_0 : P(S|A) = P(S|B) = P(S)$$

or equivalently

$$H_0 : P(F|A) = P(F|B) = P(F).$$

- The alternative hypothesis would represent heterogeneous populations, and so,

$$H_A : P(S|A) \neq P(S|B) \quad \text{or} \quad P(F|A) \neq P(F|B)$$

2×2 Tables

- Note that we could conduct such a test could using the test for proportions based on the normal distribution presented in the previous lecture.
- In this context, we might have

$$p_1 = P(S|A)$$

$$p_2 = P(S|B)$$

$$\hat{p}_1 = n_{11}/n_{1+}$$

$$\hat{p}_2 = n_{21}/n_{2+}$$

$$H_0 : p_1 = p_2 \text{ and } H_A : p_1 \neq p_2$$

2×2 Tables

- A more general test is based on the chi-square distribution.
- We will develop this test in the context of 2×2 tables, and then generalize the procedure to larger contingency tables: i.e., $r \times c$ tables (where r and/or c may be larger than 2).

2×2 Tables

Expected Counts

- The test statistic we will develop is based on comparing the observed cell counts n_{ij} to the cell counts we would expect if the null hypothesis was true.
- Let $O_{ij} = n_{ij}$ denote the observed cell count for row i and column j.
- Let E_{ij} denote the expected cell count for row i and column j.
- So how can we obtain estimates of these expected cell counts?
- First we will consider E_{11}.

2 × 2 Tables

Population	Variable: Success	Variable: Failure	Total
A	$P(S \mid A)$	$P(F \mid A)$	n_{1+}
B	$P(S \mid B)$	$P(F \mid B)$	n_{2+}
	$P(S)$	$P(F)$	
Total	n_{+1}	n_{+2}	n_{++}

$$
\begin{aligned}
E_{11} &= \text{(sample size for population A)} \times P(S|A) \\
&= \text{(sample size for population A)} \times P(S) \qquad \{\text{under } H_0\} \\
&\approx (n_{1+}) \left(\frac{n_{+1}}{n_{++}} \right) \\
&= \frac{n_{1+} n_{+1}}{n_{++}} \\
&= \frac{\text{(marginal row 1 total)} \times \text{(marginal column 1 total)}}{\text{(overall total)}}
\end{aligned}
$$

2 × 2 Tables

- In general, E_{ij} can be found as follows:

$$E_{ij} = \frac{n_{i+}\, n_{+j}}{n_{++}}$$

- Equivalently, we can write

$$E_{ij} = \frac{(\text{marginal row } i \text{ total}) \times (\text{marginal column } j \text{ total})}{(\text{overall total})}$$

2×2 Tables: Example

- The following 2×2 table displays the results of a study to investigate the effectiveness of bicycle safety helmets in preventing head injury (Thompson et al., 1989). The data consist of a random sample of 793 individuals who were involved in bicycle accidents during a specified one-year period.

Wearing Helmet	Head Injury: No	Head Injury: Yes	Total
Yes	130	17	147
No	428	218	646
Total	558	235	793

2×2 Tables: Example

Suppose we wish to test the null hypothesis of homogeneity for this table.

H_0: The proportion of individuals suffering head injuries among the population of individuals wearing safety helmets at the time of the accident is equal to the proportion of persons sustaining head injuries among those not wearing helmets.

H_A: The proportions of persons suffering head injuries are not identical in the two populations.

2×2 Tables: Example

- The expected counts would be found as follows:

$$E_{11} = \frac{147 \times 558}{793} = 103.4 \qquad E_{12} = \frac{147 \times 235}{793} = 43.6$$

$$E_{21} = \frac{646 \times 558}{793} = 454.6 \qquad E_{22} = \frac{646 \times 235}{793} = 191.4$$

2 × 2 Tables: Example

Compare the observed cell counts (top table) to the expected cell counts (bottom table):

Wearing Helmet	Head Injury: No	Head Injury: Yes
Yes	130	17
No	428	218

Wearing Helmet	Head Injury: No	Head Injury: Yes
Yes	103.4	43.6
No	454.6	191.4

2×2 Tables: Example

- Does it appear that the discrepancies between the observed cell counts and the expected cell counts are substantial enough to provide statistically significant evidence in favor of the alternative hypothesis?

2 × 2 Tables

Chi-Square Test Statistic

- The *Chi-Square Test Statistic* for testing the hypothesis of homogeneity is given by

$$X^2 = \sum_{j=1}^{2} \sum_{i=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- If the null hypothesis is true, and if all of the expected counts are at least 5, this test statistic has a *Chi-Square Distribution*.

- Like the Student's t-distribution, the shape of the density curve for the chi-square distribution depends upon degrees of freedom.

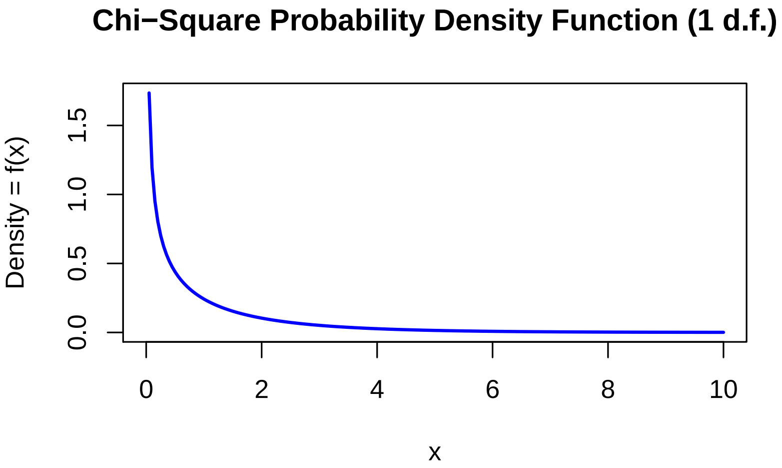
2×2 Tables

Characteristics of the chi-square density curve:

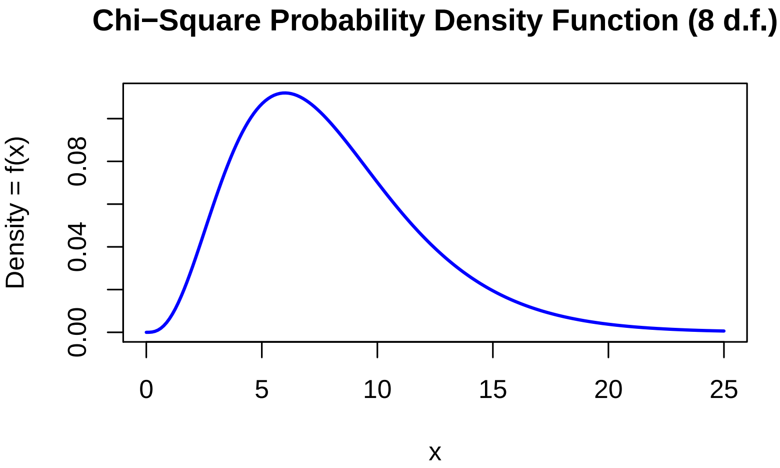
- the curve lies entirely to the right of zero (a chi-square random variable is inherently nonnegative),
- the curve is right skewed,
- as the degrees of freedom increase, the shape of the chi-square curve approaches the shape of a normal curve.

The test statistic, X^2, on the previous slide, has 1 degree of freedom.

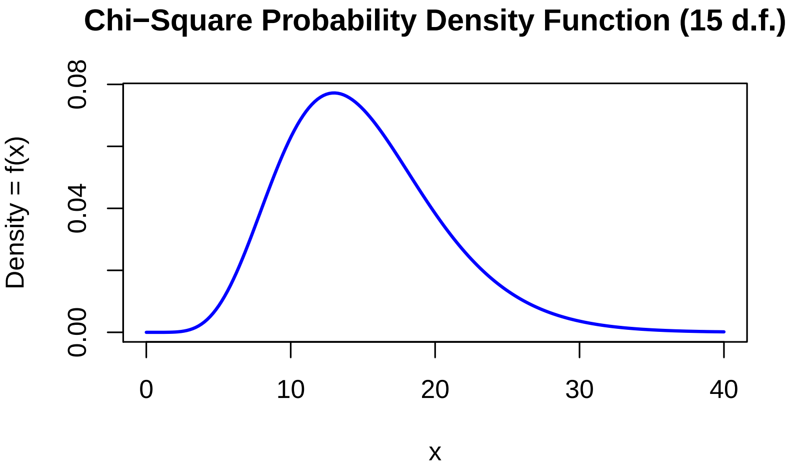
2×2 Tables



2×2 Tables



2×2 Tables



2×2 Tables

- To conduct a test of homogeneity, we will need to use critical values based on the chi-square distribution for the purpose of determining rejection regions.
- Let χ^2 denote a chi-square random variable.
- Let χ^2_α denote a cut-off point along the horizontal axis under the chi-square curve such that

$$P\left(\chi^2 > \chi^2_\alpha\right) = \alpha$$

i.e., such that the area to the right of χ^2_α is α.

2×2 Tables

- The tables supplied on ICON provide the critical values for the chi-square distribution.
- This table has the same basic layout as the one for the *t*-distribution.
- The chi-square test of homogeneity is always one-sided.
- For an α level test, we reject the null hypothesis if the test statistic, X^2, exceeds χ^2_α; i.e., if $X^2 > \chi^2_\alpha$.

2×2 Tables: Example

Returning to the bike helmet example, we have

$$
\begin{aligned}
X^2 &= \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \\
&= \frac{(130-103.4)^2}{103.4}+\frac{(17-43.6)^2}{43.6}+ \\
&\quad \frac{(428-454.6)^2}{454.6}+\frac{(218-191.4)^2}{191.4} \\
&= 6.84+16.23+1.56+3.70 \\
&= 28.33
\end{aligned}
$$

2×2 Tables: Example

- From the table, for a chi-square curve based on 1 degree of freedom, we have $\chi^2_{0.05} = 3.84$.
- Since the observed value for our test statistic exceeds this cut-off point, we would reject H_0.
- Conclusion: At the $\alpha = 0.05$ level, there is statistically significant evidence to indicate that the proportion suffering head injuries among those wearing safety helmets at the time of the accident is not equal to the proportion sustaining head injuries among those not wearing helmets.

2×2 Tables

- Note: There is a version of the test statistic based on an adjustment called Yates correction, which is negligible in large-sample settings. We will not consider this correction.

$r \times c$ Tables

- Consider drawing samples from r populations, and measuring a categorical variable with c levels on each of the subjects in the r samples.

- The resulting data could be represented in the form of a table:

	Variable			
Population	1	2	$\cdots$	c
1	n_{11}	n_{12}	$\cdots$	n_{1c}
2	n_{21}	n_{22}	$\cdots$	n_{2c}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
r	n_{r1}	n_{r2}	$\cdots$	n_{rc}

- Such a table is called an *$r \times c$ Contingency Table*.

$r \times c$ Tables

- Add row and column totals to present the marginal totals:

Population	Variable 1	2	$\cdots$	c	Total
1	n_{11}	n_{12}	$\cdots$	n_{1c}	n_{1+}
2	n_{21}	n_{22}	$\cdots$	n_{2c}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	n_{r+}
Total	n_{+1}	n_{+2}	$\cdots$	n_{+c}	n_{++}

- In the context of $r \times c$ contingency tables, we are often interested in testing the hypothesis of homogeneity: that the probability of an outcome corresponding to level j of the variable is the same for each of the r populations $(1 \leq j \leq c)$.

$r \times c$ Tables

Steps for a Chi-Square Test of Homogeneity

1. Describe the populations and the categorical variable of interest.

2. State the null hypothesis of homogeneity, H_0, and the alternative hypothesis of heterogeneity, H_A.

3. Select a value for α.

$r \times c$ Tables

4. Specify the *Test Statistic* to be used.
   For the chi-square test, the test statistic is

$$X^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

   Here, X^2 is based on $(r-1)(c-1)$ degrees of freedom (df).

$r \times c$ Tables

The expected counts E_{ij} can be found as follows:

$$E_{ij} = \frac{n_{i+}\, n_{+j}}{n_{++}}$$

Equivalently, we can write

$$E_{ij} = \frac{(\text{marginal row } i \text{ total}) \times (\text{marginal column } j \text{ total})}{(\text{overall total})}$$

Note: This test statistic can be used when the expected counts are 'large': specifically, when $E_{ij} \geq 5$.

$r \times c$ Tables

5. Compute the numerical value of the test statistic.

6. Compute the p-value for the test using a statistical software package. Alternatively, find bounds for the p-value using Table A.8.

   The p-value for the test is given by

   $$P(\chi^2 \geq X^2),$$

   where X^2 is the observed value of the test statistic (computed in step 5) and χ^2 is a chi-square random variable based on $(r-1)(c-1)$ degrees of freedom (df).

$r \times c$ Tables

7. Arrive at a conclusion by either:
   (1) comparing the p-value to α, or
   (2) determining whether the test statistic falls into the rejection region.

   The rejection region for the test is given by the set of values for the test statistic X^2 that exceed the critical value χ^2_α.

8. State the conclusion (i.e., whether or not H_0 should be rejected).

$r \times c$ Tables

Steps in Bounding a Chi-Square Test p-value Using the Table.

Let X^2 denote the observed value of the test statistic.

1. In the appropriate row, find two adjacent critical values χ^2_a and χ^2_b which bound $X^2 : \chi^2_a \leq X^2 \leq \chi^2_b$.

2. In the top row of find the upper tail areas p_1 and p_2 corresponding to χ^2_a and χ^2_b.

3. The p-value is between p_1 and p_2.

$r \times c$ Tables: p-value Bounding Example

In conducting a chi-square test based on a contingency table with $r = 2$ rows and $c = 4$ columns, suppose we obtain a test statistic of $X^2 = 8.27$. Find bounds for the p-value.

1. The degrees of freedom are
   $df = (r - 1)(c - 1) = (2 - 1)(4 - 1) = 3$.

   In the row corresponding to $df = 3$, $X^2 = 8.27$ is between $\chi^2_a = 7.81$ and $\chi^2_b = 9.35$.

2. From top row $p_1 = 0.050$ and $p_2 = 0.025$.

3. The p-value is between $p_2 = 0.025$ and $p_1 = 0.050$.

$r \times c$ Tables

χ^2 *Table*:

	Area in Upper Tail			
df	0.100	**0.050**	**0.025**	$\cdots$
1	2.71	3.84	5.02	$\cdots$
2	4.61	5.99	7.38	$\cdots$
3	6.25	**7.81**	**9.35**	$\cdots$
4	7.78	9.49	11.14	$\cdots$
5	9.24	11.07	12.83	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$r \times c$ Tables: Example

- Consider the following 2×3 table, taken from a study that investigated the accuracy of death certificates (Schottenfeld et al., 1982).
- In two different hospitals, the results of 575 autopsies were compared to the causes of death listed on the certificates.
- One hospital that participated in the study was a community hospital (A); the other was a university hospital (B).
- Conduct a test of homogeneity for the two hospitals. Use $\alpha = 0.05$.

$r \times c$ Tables: Example

	Death Certificate Status			
Hospital	Confirmed Accurate	Inaccurate; No Change	Incorrect; Recoding	Total
A	157	18	54	229
B	268	44	34	346
Total	425	62	88	575

$r \times c$ Tables: Example

Expected Cell Counts

$$E_{11} = \frac{229 \times 425}{575} = 169.3$$

$$E_{12} = \frac{229 \times \ 62}{575} = \ 24.7$$

$$E_{13} = \frac{229 \times \ 88}{575} = \ 35.0$$

$$E_{21} = \frac{346 \times 425}{575} = 255.7$$

$$E_{22} = \frac{346 \times \ 62}{575} = \ 37.3$$

$$E_{23} = \frac{346 \times \ 88}{575} = \ 53.0$$

$r \times c$ Tables: Example

$$
\begin{aligned}
X^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \\
&= \frac{(157-169.3)^2}{169.3}+\frac{(18-24.7)^2}{24.7}+\frac{(54-35.0)^2}{35.0} \\
&\quad +\frac{(268-255.7)^2}{255.7}+\frac{(44-37.3)^2}{37.3}+\frac{(34-53.0)^2}{53.0} \\
&= 0.89+1.82+10.31+0.59+6.81+1.20 \\
&= 21.62.
\end{aligned}
$$

- $P(\chi^2_2 \geq 21.62) < 0.001$.
- Since $p < 0.05$ we reject the null hypothesis of homogeneity, and conclude that the distribution of death certificate status differs between hospital A and hospital B.

$r \times c$ Tables: Example

	Death Certificate Status			
Hospital	Confirmed Accurate	Inaccurate; No Change	Incorrect; Recoding	Total
A	0.685	0.079	0.236	1.00
B	0.775	0.127	0.098	1.00

	Death Certificate Status		
Hospital	Confirmed Accurate	Inaccurate; No Change	Incorrect; Recoding
A	0.369	0.290	0.614
B	0.631	0.710	0.386
Total	1.000	1.000	1.000

McNemar's Test

- In section 11.1, we discussed the procedure for conducting a test to compare two means using paired samples.

- *McNemar's Test* is a test to compare two proportions using paired samples.

- Recall that a paired sample arises when two samples are collected in such a manner that every observation in the first sample can be matched to a corresponding observation in the second sample.

- Consider obtaining data pairs by measuring two binary ('success' / 'failure') variables of interest to obtain two samples.

McNemar's Test

- A study was done with 40 terminally ill patients to determine if they received pain relief from two different agents given three weeks apart, in random order across subjects. The two pain killers were morphine and THC (cannabis).

Morphine Pain Relief	THC Pain Relief? Yes	THC Pain Relief? No	Total
Yes	11	12	23
No	7	10	17
Total	18	22	40

- Is there sufficient evidence to claim that one of these two provides greater pain relief?

McNemar's Test

- Note that there is a total of 40 patients in this study, each measured twice.
- Some people will have the same response for each pain killer (i.e., either both provide pain relief, or both do not).
  - These are called *concordant pairs* (of observations), because the response on each is the same.
- Others will find a difference between the two pain killers—finding relief with one but not the other.
  - These are called *discordant pairs*, because the two responses of the same subject are different.
- The discordant pairs give us the most information on which drug is a more effective pain killer.

McNemar's Test

- In our example, there are 10+11 concordant pairs (diagonal counts) and $12 + 7$ discordant pairs (off-diagonals).
- McNemar's test ignores the concordant pairs.

Morphine Pain Relief	THC Pain Relief? Yes	THC Pain Relief? No	Total
Yes	11	$r = 12$	23
No	$s = 7$	10	17
Total	18	22	40

- H_0 : No preference for one pain killer over the other.
  H_A : One pain killer is preferred over the other.

McNemar's Test

- Let $r = 12$ be the number of discordant pairs where Morphine was preferred, and $s = 7$ be the number where THC was preferred.
- McNemar's test assumes $R \sim Bin(r + s, p = 1/2)$, and tests $H_0 : p = 1/2$.
- In other words, we have a Binomial data situation with the number of 'trials' (n) being $(r + s) = 19$ and $p = 1/2$.
- That is, we have the same probability of falling either above or below the diagonal.

McNemar's Test

- To get the p-value, compute the probability of getting an r (or equivalently, s) value that extreme or more extreme.
- Since $R \sim Bin(r+s, 1/2)$ is symmetric, you can just double the p-value for a two sided test.

$$
\begin{aligned}
p\text{-value} &= P(R \leq 7) + P(R \geq 12) \\
&= 2 \times P(R \leq 7) = 2 \times P(R \geq 12) \\
&= 2 \times (0.0961 + 0.0518 + 0.0222 + \ldots) \\
&= 2 \times 0.1796 = 0.3592 \\
&> 0.05 \text{ and therefore is not significant.}
\end{aligned}
$$

McNemar's Test

- In general, if we were to do a normal approximation to the binomial (with no continuity correction), we recognize that, under H_0,

$$E(R) = (r+s) \times 0.5, \text{ and}$$
$$Var(R) = (r+s) \times 0.5 \times 0.5$$

So

$$\begin{aligned} Z &= \frac{r - (r+s) \times 0.5}{\sqrt{(r+s) \times 0.5 \times 0.5}} \\ &= \frac{0.5(r-s)}{0.5\sqrt{r+s}} = \frac{r-s}{\sqrt{r+s}} \end{aligned}$$

Finally, the test statistic for McNemar's test is

$$X^2 = Z^2 = \frac{(r-s)^2}{r+s} \sim \chi^2_1.$$

McNemar's Test

- Sometimes a continuity correction of '-1' is used, (some authors use '-0.5'), resulting in the formula

$$X^2 = \frac{(|r-s|-1)^2}{r+s} \overset{\text{Under } H_0}{\sim} \chi^2_1$$

- For our example

$$X^2 = \frac{(|12-7|-1)^2}{19} = \frac{4^2}{19} = 0.842$$

The probability of a test statistic as extreme or more extreme than this is $P(\chi^2_1 > 0.842) = 0.3588$, and so we fail to find a significant difference in pain relief between the two drugs.

McNemar's Test

- Consider the general layout for n pairs of binary data:

First Sample	Second Sample: Success	Second Sample: Failure
Success	n_{11}	n_{12}
Failure	n_{21}	n_{22}

- Each person appears only once in the table, so $n_{11} + n_{12} + n_{21} + n_{22} = n$

- Let p_1 denote the proportion of successes in the population associated with the first sample, and let p_2 denote the proportion of successes in the population associated with the second sample.

McNemar's Test

- From the preceding table,

$$\hat{p}_1 = \frac{(n_{11} + n_{12})}{n} \text{ would estimate } p_1, \text{ and}$$

$$\hat{p}_2 = \frac{(n_{11} + n_{21})}{n} \text{ would estimate } p_2.$$

- McNemar's test is a test of the equality of p_1 and p_2.

McNemar's Test

Steps for McNemar's Test on Population Proportions (Paired Sample)

1. Label and describe the parameters of interest.
2. State the null hypothesis H_0 symbolically: $p_1 = p_2$.
   State the alternative hypothesis H_A symbolically: $p_1 \neq p_2$.
3. Select a value for α.

McNemar's Test

4. Specify the *Test Statistic* to be used. For McNemar's test, the test statistic is

$$X^2 = \frac{(|n_{12} - n_{21}| - 1)^2}{n_{12} + n_{21}}$$

Here X^2, is based on 1 degree of freedom (df).

Note: This test statistic can be used when $(n_{12} + n_{21})$ is 'large'. A common guideline is $(n_{12} + n_{21}) \geq 10$.

Note: The factor -1 in the numerator is a small-sample adjustment that is often discarded.

McNemar's Test

5. Compute the numerical value of the test statistic.

6. Compute the p-value for the test using a statistical software package. Alternatively, find bounds for the p-value using Table A.8.

   The p-value for the test is given by

   $$P(\chi_1^2 \geq X^2),$$

   where X^2 is the observed value of the test statistic (computed in step 5) and χ_1^2 is a chi-square random variable based on 1 degree of freedom (df).

McNemar's Test

7. Arrive at a conclusion by either:
   (1) comparing the *p*-value to α, or
   (2) determining whether the test statistic falls into the rejection region.

   The rejection region for the test is given by the set of values for the test statistic X^2 that exceed the critical value $\chi^2_{1,\alpha}$.

8. State the conclusion (i.e., whether or not H_0 should be rejected).

Measures of Association Strength

- The tests described in this chapter test whether or not an association exists, but they do not provide a measure of that association.
- This is analogous to a two-sample t-test telling us that there is a significant difference between two means, but not telling us the magnitude of that difference.
- There are many possible measures of the strength of the association, and the choice among them depends on whether the row and column variables are ordinal or nominal, and whether one of them (row or column) is intended to be used to predict the other.

Measures of Association Strength

- Commonly used measures based on the chi-square statistics are the phi coefficient, Pearson's contingency coefficient and Cramer's V, which are applicable when the measures are nominal.
- For ordinal variables, we want to know if the value of one variable tends to increase as the value of the other increases (+ association) or if it tends to decrease as the other increases (- association).
- Commonly used measures for association among ordinal variables include Gamma, Kendall's tau-b, Stuart's tau-c, and Somer's D.

The Odds Ratio

- We will focus on the **odds ratio**.
- The odds ratio is a useful measure of the strength of association between two dichotomous variables.

The Odds Ratio

- **Odds**: If an event takes place with probability p, then the odds in favor of the event are

$$\frac{p}{1-p}.$$

- E.g., if $p = \frac{2}{3}$ then the odds would be $\frac{2/3}{1-2/3} = 2$.

  Thus, the probability that the event occurs is 2 times the probability that the event does not occur.

- If we are told the odds of an event are a to b, the probability the event will occur is

$$\frac{a}{a+b}$$

The Odds Ratio

- Suppose we have the following information on two dichotomous variables representing disease and exposure.

Disease Status	Exposure Status: Exposed	Exposure Status: Unexposed	Total
Disease	n_{11}	n_{12}	n_{1+}
No Disease	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n_{++}

- The odds ratio can be defined in two mathematically equivalent ways.

The Odds Ratio

1. The odds in favor of disease among exposed individuals divided by the odds in favor of disease among unexposed

$$OR = \frac{P(D|E)/[1 - P(D|E)]}{P(D|E^c)/[1 - P(D|E^c)]} = \frac{P(D|E) \times [1 - P(D|E^c)]}{P(D|E^c) \times [1 - P(D|E)]}$$

2. The odds of exposure among diseased individuals divided by the odds of exposure among the non-diseased

$$OR = \frac{P(E|D)/[1 - P(E|D)]}{P(E|D^c)/[1 - P(E|D^c)]} = \frac{P(E|D) \times [1 - P(E|D^c)]}{P(E|D^c) \times [1 - P(E|D)]}$$

The Odds Ratio

- We can estimate the odds ratio by the cross product ratio

$$\widehat{OR} = \frac{n_{11}\, n_{22}}{n_{12}\, n_{21}}$$

- Similarly, the estimator of the relative risk (the ratio of the probabilities of disease comparing exposed to unexposed) is

$$\widehat{RR} = \frac{n_{11}\,(n_{12} + n_{22})}{n_{12}\,(n_{11} + n_{21})}$$

- If a disease is rare, thus the values of n_{11} and n_{12} are small relative to the values of n_{21} and n_{22}, then the relative risk can be approximated by the odds ratio.

The Odds Ratio

- To better quantify the uncertainty of $\widehat{OR}$ we would like to construct a confidence interval.
- Since the odds ratio is always positive, we will first find a confidence interval for the log odds ratio.
- It turns out that the distribution of $\ln(\widehat{OR})$ is more symmetric and is approximately normal.
- The estimated standard error of the log odds ratio is given by

$$\widehat{SE}[\ln(\widehat{OR})] = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

The Odds Ratio

- If any cell in the table is equal to zero, then the standard error estimate is modified to be

$$\widehat{SE}[\ln(\widehat{OR})] = \sqrt{\frac{1}{n_{11}+0.5} + \frac{1}{n_{12}+0.5} + \frac{1}{n_{21}+0.5} + \frac{1}{n_{22}+0.5}}$$

The Odds Ratio

- A $100(1-\alpha)\%$ confidence interval for the log of the odds ratio is then calculated in the usual way. That is,

$$\ln(\widehat{OR}) \pm z_{\alpha/2}\,\widehat{SE}[\ln(\widehat{OR})]$$

- To get the confidence interval for the odds ratio itself we exponentiate the confidence limits of the log odds ratio, resulting in

$$\left(e^{\ln(\widehat{OR})-z_{\alpha/2}\,\widehat{SE}[\ln(\widehat{OR})]},\;\; e^{\ln(\widehat{OR})+z_{\alpha/2}\,\widehat{SE}[\ln(\widehat{OR})]}\right)$$

The Odds Ratio: Example

- Suppose we wish to study where there is an increased risk of cesarean section when electronic fetal monitoring is used during labor. A study of 5,824 infants is conducted and summarized by

Cesarean Delivery	EFM Exposure: Yes	EFM Exposure: No	Total
Yes	358	229	587
No	2492	2745	5237
Total	2850	2974	5824

The Odds Ratio: Example

- What are the odds of C-Section in the monitored group relative to those that are not monitored?

$$\widehat{OR} = \frac{n_{11}\, n_{22}}{n_{12}\, n_{21}} = \frac{358 \times 2745}{229 \times 2492} = 1.72$$

- The odds of being delivered by C-section are 1.72 times higher for fetuses monitored by EFM during labor than those not monitored.

- This is a moderate association, but it does **not** imply that EFM monitoring causes C-section.

The Odds Ratio: Example

- What is the associated 95% Confidence Interval for this point estimate?

$$\ln(\widehat{OR}) = \ln(1.72) = 0.542$$

and

$$\widehat{SE}[\ln(\widehat{OR})] = \sqrt{\frac{1}{358} + \frac{1}{229} + \frac{1}{2492} + \frac{1}{2745}} = 0.089$$

Thus a 95% CI of the log odds ratio is

$$0.542 \pm 1.96 \times 0.089 = (0.368, \ \ 0.716)$$

The Odds Ratio: Example

- So the 95% CI for the odds ratio is

$$\left(e^{0.368}, e^{0.716}\right) = (1.44, 2.05)$$

- We are 95% confident that the odds of delivery by C-section are between 1.44 and 2.05 times higher for fetuses monitored by EFM than those not monitored.

- If this interval contained 1, it would imply that fetuses monitored versus those not monitored have identical odds of C-section.