Chi-square Distribution

If Z has a standard normal distribution, then Z^2 has a chi-square distribution, denoted by, χ^2 ,with 1 df.

Thus

$$P(|Z| > 1.96)=0.05$$

implies that

$$P(\chi^2 > 3.84)=0.05$$

because $(1.96)^2 = 3.84$.

- A Chi-square random variable is nothing more than a standard normal RV squared.
- Furthermore if $Z_1,\ldots,Z_n$ ~i.i.d. N(0,1) then

$$Z^2_1+\ldots+Z^2_n = \chi^2 \text{ is chi-square with n df.}$$

Then general form of the Chi-square statistic is:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i's are observed counts, E_i's are expected counts under the Null hypothesis, and *i* denotes the number of 'cells'.

The most common use of the Chi-square distribution is with contingency tables. Take the familiar 2x2 table as an example.

Observed table or O_i's for $i=1,\ldots,4$

	Sample 1	Sample 2	
Success	a	b	a+b
Failure	c	d	c+d
	a+c	b+d	N

where $\hat{\theta}_1 = \frac{a}{a+c}$ *and* $\hat{\theta}_2 = \frac{b}{b+d}$
(it doesn't matter if we use row or column probabilities)

Under the Null hypothesis that H_0: $\theta_1=\theta_2$, which implies that the rows and columns are independent, we have

Expected table or E_i's for $i=1,\ldots,4$

	Sample 1	Sample 2	
Success	(a+b)(a+c)/N	(a+b)(b+d)/N	a+b
Failure	(c+d)(a+c)/N	(c+d)(b+d)/N	c+d
	a+c	b+d	N

(Why?)

Our test statistic is the standard observed - expected chi-square statistic

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \left[\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\theta_p(1-\theta_p)\left(\frac{1}{a+c} + \frac{1}{b+d}\right)}} \right]^2$$

$$\hat{\theta}_1 = \frac{a}{a+c} \quad and \quad \hat{\theta}_2 = \frac{b}{b+d} \qquad \theta_p = \frac{a+b}{a+b+c+d}$$

which is simply the Z-statistic for the difference between the two proportions squared. Because the Z-statistic is only approximately normal in <u>large</u> samples, the χ^2 will only be approximately correct in large samples as well.

Notice that we are testing that conditional probabilities of success given the column (or sample) are equal.

The χ^2 test statistic is almost always used with a 2x2 table, but the hypotheses being tested changes depending on how we obtained the data.

<u>Case 1</u>: Prospective or Cross Sectional study

	Exposed	Not Exposed	
Disease	a	b	a+b
No Disease	c	d	c+d
	a+c	b+d	N

where $\hat{\theta}_1 = \frac{a}{a+c}$ *and* $\hat{\theta}_2 = \frac{b}{b+d}$

θ_1 =P(disease| exposed)

θ_2 =P(disease| not exposed)

H_0: $\theta_1=\theta_2$ or H_0: rr $=\theta_1/\theta_2=1$

or

H_0: or $=\theta_1/(1-\theta_1)/\theta_2/(1-\theta_2)=1$

The null hypothesis is often interpreted as "No association between Exposure and Disease" or "The rows and columns are independent", i.e. P(D|E)=P(D| not E) implies that disease status is independent of event E.

Case 2: Retrospective study (Case-control)

	Exposed	Not Exposed	
Disease	a	b	a+b
No Disease	c	d	c+d
	a+c	b+d	N

where $\hat{\theta}_1 = \frac{a}{a+b}$ *and* $\hat{\theta}_2 = \frac{c}{c+d}$

θ_1 =P(exposed | disease)
θ_2 =P(exposed | no disease)

- Because of our sampling scheme a+c is fixed, and thus we can never estimate the true relative risk. Hence

H_0: or $=\theta_1/(1-\theta_1)/\theta_2/(1-\theta_2)=1$

Remember:

$$or = \frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)} = \frac{P(D|E)}{P(D|\overline{E})} \times \left[\frac{1-P(D|\overline{E})}{1-P(D|E)}\right] = rr \times \left[\frac{1-P(D|\overline{E})}{1-P(D|E)}\right]$$

For contingency tables with r rows and c columns, use the same chi-square test statistic:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad \text{df=(r-1)(c-1)}$$

Where the DF=(r-1)(c-1). This is often referred to as an 'test of association'. That is we are testing to seem if the rows and columns are associated or if they are `independent'.

Statistically we are testing to see if **<u>all</u>** of row probabilities, conditional on the columns, are equal. For example:

	1	2	3	4	5
A	θ_{a1}	θ_{a2}	θ_{a3}	θ_{a4}	θ_{a5}
B		θ_{b2}			θ_{b5}
C			θ_{c3}		θ_{c5}
D				θ_{d4}	θ_{d5}
E					θ_{e5}

The Chi-square test of no association has the null

H_0: $\theta_{a1}=\theta_{a2}=\theta_{a3}=\theta_{a4}=\theta_{a5}$ and
$\theta_{b1}=\theta_{b2}=\theta_{b3}=\theta_{b4}=\theta_{b5}$ and
$\theta_{c1}=\theta_{c2}=\theta_{c3}=\theta_{c4}=\theta_{c5}$ and
$\theta_{d1}=\theta_{d2}=\theta_{d3}=\theta_{d4}=\theta_{d5}$ and
$\theta_{e1}=\theta_{e2}=\theta_{e3}=\theta_{e4}=\theta_{e5}$ and

Sometimes we wish to combine several contingency tables. When this happens we have to be careful. One way of determining if the data are similar enough to combine is to examine their odds ratio.

	Sample 1	Sample 2	
Success	a	b	a+b
Failure	c	d	c+d
	a+c	b+d	N

where $\hat{\theta}_1 = \dfrac{a}{a+c}$ *and* $\hat{\theta}_2 = \dfrac{b}{b+d}$

In this table the estimated or is

$$\theta_1/(1-\theta_1)/\theta_2/(1-\theta_2)=ad/bc$$

and the confidence interval for the *LOG OR* is:

$$\log(o\hat{r}) \pm Z_{\alpha/2}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

$$\log(o\hat{r}) \pm Z_{\alpha/2}\sqrt{\frac{1}{\hat{\theta}_1(a+c)}+\frac{1}{(1-\hat{\theta}_1)(a+c)}+\frac{1}{\hat{\theta}_2(b+d)}+\frac{1}{(1-\hat{\theta}_2)(b+d)}}$$

⇨ Why the Log?

There is a test called the Mantel-Haenszel test, which tests if the odds ratio is equal in different r x c contingency tables. This test is often called the 'test of Homogeneity'.

We will discuss it in detail later, but now we will examine a case when two 2x2 tables are combined.

Example -- Simpson's Paradox

This example is from an investigation of the relationship between smoking and cancer (data are fictitious). Gender is associated with both of these variables.

Here we will see that the odds of having cancer is greater for smokers, for both females and males.

However when we combine the data we see that smoking is actually protective for cancer. (why ?)

Males	Smoking		
cancer	Yes	No	
Yes	**9**	**51**	60
No	**6**	**43**	49
	15	94	109

Or(males)=(9x43)/(6x51)=1.26

P(C|S) = 9/15=0.6
P(C|nS) = 51/94=0.543
RR= 1.105; Risk Diff= 0.057

Females	Smoking		
cancer	Yes	No	
Yes	**14**	**7**	21
No	**19**	**12**	31
	33	19	52

Or(females)=(140x120)/(190x70)=1.26

P(C|S) = 14/33=0.42
P(C|nS) = 7/19=0.368
RR= 1.14; Risk Diff= 0.052

Both the male and the female have the same trend: the odds of developing cancer are higher among smokers than non-smokers.

Since the estimated odds ratios are similar why not simply combine the two tables?

combined	Smoking		
cancer	Yes	No	
Yes	**23**	**58**	81
No	**25**	**55**	80
	48	113	161

Or(combined)=(23x55)/(58x25)=0.872=1/1.14

P(C|S)   = 23/48=0.479
P(C|nS)   = 58/113=0.513
RR= 0.934=1/1.071 ; Risk Diff= -0.034

And we see that smoking is now protective! Why did does the trend reverse?

- What does this imply about simply testing for homogeneity and then combining over tables?

⇨ Be very, very careful.

Sometimes, it is better to present stratified data when the effect appears the same, than it is to present the combined data.

Another example:
Ask Marilyn (Parade Magazine, April 28, 1996)

A reader writes in to ask:

"A company decided to expand, so it opened a factory generating 455 jobs. For the 70 white collar positions, 200 males and 200 females applied. Of the females who applied, 20% were hired, while only 15% of the males were hired. Of the 400 males applying for the blue collar positions, 75% were hired, while 85% of the females were hired.

A federal Equal Employment enforcement official noted that many more males were hired than females, and decided to investigate. Responding to charges of irregularities in hiring, the company president denied any discrimination, pointing out that in both the white collar and blue collar fields, the percentage of female applicants hired was greater than it was for males.

But the government official produced his own statistics, which showed that a female applying for a job had a 58% chance of being denied employment while male applicants had only a 45% denial rate. As the current law is written, this constituted a violation....

Can you explain how two opposing statistical outcomes are reached from the same raw data?"

Here is the data:

White	Hired		
	Yes	No	
Male	**30**	**170**	200
Female	**40**	**160**	200
	70	330	400

P(H|M) = 30/200=0.15
P(H|F) = 40/200=0.20

Blue	Hired		
	Yes	No	
Male	**300**	**100**	400
Female	**85**	**15**	100
	385	115	500

P(H|M) = 300/400=0.75
P(H|F) = 85/100=0.85

All	Hired		
	Yes	No	
Male	**330**	**270**	600
Female	**125**	**175**	300
	455	445	900

P(H|M) = 330/600=0.55
P(H|F) = 125/300=0.42

Marilyn, correctly notes that, even though all the figures presented are correct, the two outcomes are not opposing:

"Say a company tests two treatments for an illness. In trial No. 1, treatment A cures 20% of its cases (40 out of 200) and treatment B cures 15% of its cases (30 out of 200). In trial No. 2, treatment A cures 85% of its cases (85 out of 100) and treatment B cures 75% of its cases (300 out of 400)....

So, in two trials, treatment A scored 20% and 85%. Also in two trials, treatment B scored only 15% and 75%. No matter how many people were in those trials, treatment A (at 20% and 85%) is surely better than treatment B (at 15% and 75%), right?
Wrong! Treatment B performed better. It cured 330 (300+30) out of the 600 cases (200+400) in which it was tried --a success rate of 55%...By contrast, treatment A cured 125 (40+85) out of the 300 cases (200+100) in which it was tried, a success rate of only about 42%."

She notes that this is exactly what happened to the employer. Because so many more men applied for the blue collar positions, even if the employer hired all the women who had applied for blue collar positions, it couldn't satisfy the government regulations.

Exact Tests for Probabilities

To estimate the expected value of a random variable on the basis of n independent observations, we use the sample mean. To set a confidence interval, we can use the fact that

$$\frac{\overline{X}_n - \mu}{\hat{S}E(\overline{X}_n)}$$

has an approximate standard normal distribution. (The bigger n is, the better the approximation.) From this we find, when n is large,

95% Confidence Interval: $\overline{X}_n \pm 1.96\ \hat{S}E(\overline{X}_n)$

Level α= 0.05 Test: Reject $H_0: \mu = \mu_0$ vs $H_A: \mu > \mu_0$ if

$$\frac{\overline{X}_n - \mu_0}{\hat{S}E(\overline{X}_n)} > 1.645 .$$

If when the hypothesis $H_0: \mu = \mu_0$ is true, the value of $SE(\overline{X}_n)$ is known, then that value often replaces the estimate, $\hat{S}E(\overline{X}_n)$, in the denominator of the test statistic.

For example, this is the case when the *X*'s have a Bernoulli(θ) distribution, because $H_0: \theta = \theta_0$ implies that $SE(\overline{X}_n) = \sqrt{\theta_0(1-\theta_0)/n}$. The usual test statistic is $(\overline{X}_n - \theta_0)/\sqrt{\theta_0(1-\theta_0)/n}$.

To test the hypothesis that the probability of heads with my 40¢ piece is 0.50 (just like most other coins) we toss it 80 times, and observe the number of heads. Suppose we see 50 heads in our 80 tosses.

Under the null hypothesis, H_0, we <u>know</u> the variance. It is $\theta_0(1-\theta_0)$, which in the present example, is 1/4. So instead of <u>estimating</u> the variance, we can use the test statistic,

$$\frac{\sqrt{n}\,(\hat{\theta} - 0.5)}{\sqrt{0.25}},$$

rejecting the hypothesis $H_0: \theta = 0.5$ vs $H_A: \theta > 0.5$ if the test statistic exceeds the critical value found in the z-table.

For our observations (50 heads in 80 tosses),
$\hat{\theta} = 50/80 = 0.625$,
so the observed value of the approximate normal test statistic is

$$\frac{\sqrt{80}(0.625-0.5)}{\sqrt{0.25}} = 2.236$$

and the approximate p-value is $P(Z \geq 2.236) = 0.013$.This evidence is strong enough to justify rejecting H_0 at the 5% level, but not at the 1% level (0.01< p < 0.05).

The general procedure is this: Reject $H_0 : \theta = \theta_0$ if the approximate p-value is $\leq \alpha$:

$$P\left(Z > \frac{\sqrt{n}(\hat{\theta}_{obs} - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right) < \alpha$$

This test is based on the fact that if H_0 is true,

$$E(\hat{\theta}) = \theta_0 \quad Var(\hat{\theta}) = \theta_0(1-\theta_0)/n$$

and if n is large, the probability distribution of $\hat{\theta}$ can be approximated by the normal distribution with this mean and variance.

STATA does this test for me at the command **prtest** which instructs it to perform "One- and two-sample tests of proportions":

```
. prtesti 80 50 0.5

One-sample test of proportion                      x: Number of obs =        80

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       .625    .0541266    11.547   0.0000       .5189138    .7310862
------------------------------------------------------------------------------

                          Ho: proportion(x) = .5

       Ha: x < .5              Ha: x ~= .5              Ha: x > .5
        z =  2.236              z =  2.236              z =  2.236
      P < z = 0.9873         P > |z| = 0.0253         P > z = 0.0127
```

But if H_0 is true, we know not only the mean and the variance; we know the entire probability distribution of $\hat{\theta}$. Since $\hat{\theta} = X/n$, where $X \sim binomial(n,\theta_0)$, we know that the probability that $\hat{\theta}$ will equal k/n is precisely

$$\binom{n}{k}\theta_0^k(1-\theta_0)^{n-k} \qquad for \quad k = 0,1,...,n$$

Thus we can use the exact probability distribution—we don't have to use an approximation.

For the 80 tosses of my 40¢ piece, with $H_0: \theta = 0.5$ and $H_A: \theta > 0.5$ the exact p-value is

$$p\text{-}value = P(\hat{\theta} \geq 0.625) = P(X \geq 50 \mid \theta = 0.5)$$

$$= \sum_{x=50}^{80} \binom{80}{x} (0.5)^x (1 - 0.5)^{80-x}$$

$$= 0.016496$$

STATA does this test for me at the command **bitest** which instructs it to perform a "Binomial probability test":

```
. bitesti 80 50 0.5

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       80         50           40       0.50000      0.62500

  Pr(k >= 50)            = 0.016496  (one-sided test)
  Pr(k <= 50)            = 0.990842  (one-sided test)
  Pr(k <= 30 or k >= 50) = 0.032993  (two-sided test)
```

The real usefulness of the exact test procedure appears when the sample size is small. We can perform precise hypothesis tests, even when we have only a small sample.

In general, to test $H_0: \theta = \theta_0$ *vs* $H_A: \theta > \theta_0$ *at level* $\alpha = 0.05$ *we will reject* H_0 *if* $\hat{\theta} = X/n$ *is too large, i.e., if* X *is too large. How large? Instead of rejecting if the approximate p-value is* ≤ 0.05 *, calculate the* <u>*exact*</u> *p-value and reject* H_0 *if it is* ≤ 0.05*:*

$$p\text{-}value = P(X \geq x_{obs}) = \sum_{x_{obs}}^{n} \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k}$$

Example: Test $H_0: \theta = 1/3$ vs $H_A: \theta > 1/3$. If n = 4 and we observe 3 successes, what is the p-value?

$$P(\ X\ \geq 3\ |\ \theta = 1/3\ ) = \binom{4}{3}(1/3\ )^3(\ 2/3\ ) + (\ 1/3\ )^4 = \frac{1}{9} = 0.11111\ .$$

If n=4 and we observe 4 successes, what is the p-value?

$$P(\ X\ \geq\ 4\ |\ \theta = 1/3\ ) = (\ 1/3\ )^4 = \frac{1}{81} = 0.012346\ .$$

Suppose we want a <u>two-sided</u> test of
$H_0: \theta = 1/3$ vs $H_A: \theta \neq 1/3$
when n = 4 and we observe 4 successes. What is the p-value?

One solution: Calculate the one-sided test p-value and double it. This gives p = 0.02469. But this procedure can lead to p-values > 1. For example, for testing $H_0: \theta = 1/3$ vs $H_A: \theta > 1/3$ when n = 5, and x = 2 is observed, the (one-sided) p-value is

$$p\text{-}value = P(\ \hat{\theta} \geq 0.40) = P(\ X \geq 2\ |\ \theta = 1/3\ ) = 0.53909$$

Doubling this gives a two-sided "p-value" of 1.078

Other suggestions have been made, and the previous release of STATA reported one of them, as well as the simple "double the one-sided p-value" procedure. But they also have their unsatisfactory aspects. Statistics has no definitive solutions for these very simple problems (What is the "right way" to do exact two-sided tests of hypotheses about binomial probabilities, and what is the "right way" to set exact two-sided confidence intervals?).

An important advantage of the "approximate normal" testing and confidence interval machinery is that in the case of the nice symmetric normal distribution it is clear what to do about two-sided tests and confidence intervals.

What to do when you face this problem?

(a) Follow whatever conventions you find in your department/discipline.
(b) Report one-sided p-value, clearly identified.

Most readers will simply double it if they "want to be conservative."

Exercise: (For education and entertainment only) Suppose you and your friend pass the time by playing a game (instead of studying Biostatistics). Your friend insists that it is "all luck," but you think the game requires some skill, of which you are confident that you have more than your friend. Suppose you keep score and find that you have won 8 games out of ten.

(a) What probability model might you use to test the hypothesis

$$H_0: \textit{All luck} \quad vs \quad H_A: \textit{You are better player} \quad ?$$

(b) What is the p-value?

Fisher's Exact Test of Equality of Two Probabilities

We have seen three ways to test the hypothesis that two binomial probabilities are equal:

(i) calculate

$$\frac{|\hat{\theta}_1 - \hat{\theta}_2|}{\sqrt{\hat{\theta}_p(1-\hat{\theta}_p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

(approximate) $p\text{-}value = P(\,|Z| \geq |z_{obs}|\,)$, and

(ii) calculate

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

(approximate) $p\text{-}value = P(\,\chi^2 \geq \chi^2_{obs}\,)$.

These two tests are actually equivalent. When the numbers of observations in the four cells are not very large, a "continuity correction" is often made in an effort to improve the accuracy of the approximate p-value or Type I error rate.

The correction reduces the value of the test statistic, thereby increasing the p-value, and making it more difficult to reject the null hypothesis:

(iii) calculate

$$\chi_c^2 = \sum \frac{(|observed - expected| - 0.5)^2}{expected}$$

(approximate) $p\text{-}value = P(\chi^2 \geq \chi^2_{c\ obs})$.

Exercise: If I perform a chi-square test on a 2-by-2 table without the continuity correction, then later do the test with the correction, how will the second p-value (the corrected one) compare to the first?

- ☐ It will definitely be smaller.
- ☐ It will be the same.
- ☐ It might be larger or smaller, depending on the numbers in the particular table that I am analyzing.
- ☐ It will definitely be larger.

Another solution was invented by the greatest of the statisticians, Sir Ronald A. Fisher. It is called "Fisher's Exact Test," and provides exact p-values no matter how small the numbers in the table.

	Success	Failure	
Sample 1	a	b	$a+b=n_1$
Sample 2	c	d	$c+d=n_2$
	a+c	b+d	n

This test is based on recognizing an important fact about the probabilities of tables of this sort: if the success probabilities in the two rows are equal (H_0), then given the total number of successes, $a+c$, the conditional probability that a of the successes will occur in the first row (and c in the second) is

$$\frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{n}{a+b}}$$

It is as if we simply had *n* subjects of which *a+c* are "successes" and chose $n_1 = a + b$ of them at random to be "sample 1," letting the rest be "sample 2." The above expression gives the probability that, if we divide the *n* subjects in this way, we will find that *a* of the *a+c* successes are in "sample 1."

	Success	Failure	
Sample 1	8	2	$10=n_1$
Sample 2	1	4	$5=n_2$
	9	6	15=n

If sample 1 and sample 2 really represent, in effect, a random division of the 15 subjects into groups of 10 and 5, what is the probability that 8 of the 9 successes will go into sample 1?

$$\frac{\binom{9}{8}\binom{6}{2}}{\binom{15}{10}} = \frac{135}{3003} = 0.045$$

Fisher's exact test calculates as the p-value the probability of the observed table, plus the probabilities of all tables that are "as extreme or more so" than the one observed.

For a one-sided test "More extreme" tables are those with the same margins as the observed table and either

(i) even larger values of "*a*" than the one observed (when the observed "*a*" is greater than the value expected under H_0, $(a+b)(a+c)/n$, as it is in our example), or

(ii) even smaller values of "*a*" than the one observed (when the observed "*a*" is smaller than the value expected under H_0).

For a two-sided test "More extreme" tables are those with the same margins as the observed table, and which are "more improbable" under H_0 (regardless of whether "*a*" is in the same side of the expected value as the one observed or not).

In our example, the observed *a*=8 is larger than the value expected under H_0, 10(9/15)=6, so the table

	Success	Failure	
Sample 1	9	1	$10=n_1$
Sample 2	0	5	$5=n_2$
	9	6	15=n

is more extreme in the same direction, with probability

$$\frac{\binom{9}{9}\binom{6}{1}}{\binom{15}{10}} = \frac{6}{3003} = 0.002$$

Since there are no more extreme tables in this direction (no tables with the same margins and even larger values of "*a*" than 9), the one-sided p-value is 0.045 + 0.002 = 0.047.

To calculate the two-sided p-value, we must find the tables (with the same margins) that are as improbable as the one observed (or more so) in the other direction (few successes in sample 1, i.e., small values of "*a*"). The first candidate is

	Success	Failure	
Sample 1	4	6	$10=n_1$
Sample 2	5	0	$5=n_2$
	9	6	$15=n$

$$\frac{\binom{9}{4}\binom{6}{6}}{\binom{15}{10}} = \frac{126}{3003} = 0.042$$

	Success	Failure	
Sample 1	5	5	$10=n_1$
Sample 2	4	1	$5=n_2$
	9	6	$15=n$

The probability of this table under H_0 (conditional probability, given a total of 9 successes) is greater than that of the observed table:

$$\frac{\binom{9}{5}\binom{6}{5}}{\binom{15}{10}} = \frac{756}{3003} = 0.252$$

so this last table's probability is <u>not</u> included in the two-sided p-value, which is the probability of the observed table (with *a* = 8) plus the probabilities of the two more extreme tables (with *a* = 9 and *a* = 4): 0.045 + 0.002 + 0.042 = 0.089. (Note that *a* = 3 is even further from the expected value than *a* = 4, but it is not possible, given the marginal totals of the table.)

STATA does the calculation for you:

```
. tabi 8 2 \ 1 4

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |         8          2 |        10
         2 |         1          4 |         5
-----------+----------------------+----------
     Total |         9          6 |        15

           Fisher's exact =                 0.089
   1-sided Fisher's exact =                 0.047
```