Chi-Square

• Parametric statistics, such as *r* and t, rest on estimates of population parameters (x for μ and s for σ) and require assumptions about population distributions (in most cases normality) for their probability calculations to be correct.
• Sometimes these assumptions cannot be met.
• The branch of statistics that concerns testing without estimating parameters is called nonparametric statistics.
•The most popular, and commonly used, approach of nonparametrics is called chi-square (χ^2).
• Our use of the test will always involve testing hypotheses about frequencies (although χ^2 has other uses).
• The two main uses of chi-square are called **goodness-of-fit** and **test for independence**.
• Overall, the Pearson χ^2 statistic measures the variability in categorical data between the actual observed frequency (O) in the different categories and the expected (predicted) frequency (E) in each category.
• When these differences are greater, χ^2 is larger.

The **goodness-of-fit test** involves a single (1) independent variable.
The **test for independence** involves 2 or more independent variables.

Assumptions for χ^2

1. χ^2 works if you have at least 5 counts in each cell.
2. However, all counts ≥ 1 and most (> 75%) of the counts should be ≥ 5.
3. Use chi-square with nominal and discrete-level data.
4. Need independent observations.

Goodness-of-Fit Test

Example: Suppose we do a blind coffee tasting experiment. We present 4 brands of coffee in identical glasses. After sampling all 4, each judge selects his or her favorite coffee. We select 100 lucky participants for the study. Alpha = .05.

Coffee	Starbucks	Maxwell House	Folgers	Seattle's Best
Frequency	10	45	10	35

If people cannot tell the difference, or if they have no preference among coffee, we would expect about 25 people to choose each coffee type.

H_0: In the population, there is no preference for any specific coffee type
H_1: In the population, one or more coffee types are preferred over the others

There are 4 choices and 100 people: 100/4 = 25.

To test, we use the formula

$$x^2 = \sum \frac{(O-E)^2}{E}$$

where O is an observed score and E is an expected score in a cell.

In our case, we expect 25 people per cell (coffee).

Coffee	Starbucks	Maxwell house	Folgers	Seattle's Best
Frequency (O)	10	45	10	35
Expected (E)	25	25	25	25
O-E	-15	20	-15	10
$O-E^2$	225	400	225	100
$O-E^2/E$	9	16	9	4

$\chi^2 = 38$

The df for this test are k-1, where k is the number of groups or cells. For this example, k = 4 and df = 3.

If we look up the values in the chi square table, we find for 3 df:

χ^2 (crit) $\alpha_{.05} = 7.81$ and χ^2 (crit) $\alpha_{.01} = 11.34$

Because our obtained value χ^2 is 38, and is beyond the critical region at the .05 level of 7.81, we can reject the null and conclude that the 4 types of coffee are not likely to be equally preferred. That is, there are differences among the 4 types of coffee, with some selected more often than others and others selected less than would be expected by chance.

NOTE:

In SPSS, the **residuals** are based on the difference between the observed (O) and the expected (E) values. The **unstandardized residual** is the simple difference of the observed and expected values.

Unstandardized residual = O - E

The **standardized residual** is found by dividing the difference of the observed and expected values by the square root of the expected value.

Standardized residual = O - E / √E

The standardized residual can be interpreted as any standard score. The mean of the standardized residual is 0 and the standard deviation is 1. Standardized residuals are calculated for each cell in the design. They are useful in helping to interpret chi-square tables by providing information about which cells contribute to a significant chi-square.

If the standardized residual is beyond the range of ± 2, then that cell can be considered to be a major contributor, if it is > +2, or a very weak contributor, if it is beyond -2, to the overall chi-square value.

The **adjusted standardized residuals** are standardized residuals that are adjusted for the row and column totals. The adjusted standardized residual is defined as:

Adjusted standardized residual = O - E / SQRT[nA * nB * (1 - nA/N) * (1 - nB/N) / N]

nA is the row total, nB is the column total, and N is the total number of cases.

Test for Independence

Example: Suppose we are interested in attitudes, but we are a newspaper with a limited amount of time and expertise in attitudinal surveys. So we decide to do an exit survey of people voting in DeKalb County. We want to know male and female attitudes about certain issues, so we ask each person exiting the poll to tell us how they voted on three measures. Alpha = .05.

	Tax increase for schools	Ban EEO hiring preferences	Tax increase for police	Total
Male	40	65	55	160
Female	70	50	60	180
Total	110	115	115	340

H_0: In the population, there is no association between the distribution of attitudinal preferences for men and the distribution of attitudinal preferences for women (i.e., the distributions have the same shape or proportions)

H_1: In the population, there is an association between the distribution of attitudinal preferences for men and the distribution of attitudinal preferences for women (i.e., the distributions do not have the same shape or proportions)

If there is no association, we would expect each cell to be the percentage of the total people, adjusted for the frequency of that row and column.

To find the expected number of people in each cell, we can use:

E = (row total x column total) / grand total

Expected values

	Tax increase for schools	Ban EEO hiring preferences	Tax increase for police	Total
Male	(110*160)/ 340= 51.76	(115*160)/ 340= 54.12	(115*160)/ 340= 54.12	160
Female	(110*180)/ 340= 58.24	(115*180)/ 340= 60.88	(115*180)/ 340= 60.88	180
Total	110	115	115	340

$(O\text{-}E)^2$ / E values

	Tax increase for schools	Ban EEO hiring preferences	Tax increase for police	Total
Male	$(40-51.76)^2/$ 51.76 = 2.67	$(65-54.12)^2/$ 54.12 = 2.19	$(55-54.12)^2/$ 54.12 = .02	
Female	$(70-58.24)^2/$ 58.24 = 2.37	$(50-60.88)^2/$ 60.88 = 1.94	$(60-60.88)^2/$ 60.88 = .01	
Total	= 5.04 +	= 4.13 +	= .03	χ^2 = **9.20**

The df for this chi-square are (row - 1) x (column - 1)

Or (2-1) x (3-1) or 2.

From our table, we find that the critical values are:
χ^2 (crit) $\alpha_{.05}$ = 5.99 and χ^2 (crit) $\alpha_{.01}$ = 9.21.

Since we stated a χ^2 (crit) α =.05 with 2 df, the results are statistically significant and we reject the H_0 because the obtained chi-square value (i.e., 9.20) exceeded the critical value of 5.99. That is, there was a statistically significant association between the distribution of choices pertaining to tax increases for schools, for police, and the banning of hiring preferences for males versus females in DeKalb county.

Confidence Intervals for 2x2 Tables:

$$(P_1 - P_2) \pm Z_{\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

Males P_1 = .333; n_1 = 120 and Females P_2= .276; n_2 = 152

1-.333 = .667 x .333 = .222
.222/120 = .00185

1-.276 = .724 x .276 = .200
.200/152 = .00132

.00185 + .00132 = .00317
SQRT (.00317) = .056 or the standard error

$P_1 - P_2$ = .333 - .276 = .057

.057 ± 1.96 x .056
The 95% CI is (-.053, .166).

All of the subsequent effect sizes (ES) can be found in SPSS.

An effect size for a larger than 2 x 2 table is Cramér's V for nominal x nominal variables, where: V = SQRT (χ^2 / (n * df))
= 9.20 / 340(2)
= 9.20 / 680 = .0135
= SQRT (.0135) = **.12**
Thus, there is a small relationship between gender and attitudinal preferences.

For V:
1 df = <.30 is a small ES
1 df = .30 to < .50 is a moderate ES
1 df = > .50 is a large ES
2 df = <.21 is a small ES
2 df = .21 to < .35 is a moderate ES
2 df = > .35 is a large ES
3 df = <.17 is a small ES
3 df = .17 to < .29 is a moderate ES
3 df = > .29 is a large ES

Note 1: For a 2 x 2 table, with nominal x nominal variables, we can use the Phi ES, where: Φ = SQRT (χ^2 / n)
Note 2: For nominal x ordinal variables, use either Phi or Cramér's V for the ES.
Note 3: For ordinal x ordinal variables, use Gamma for the ES.

Yates' Correction for Continuity:

• For a χ^2 with 1 df, a 2 x 2 table, and has expected frequencies < 5, a correction measure can be used because of this broken assumption.
• This correction measure should **only** be used in this, exact situation.
• Yates' Correction has been noted to be overly conservative in its χ^2 value and, thus, some do not recommend its use.

Example:

	Male	**Female**
Female	3	5
Male	10	2

a x d = 6
b x c = 50
6-50 = -44^2 X 20 = 38720
a + b = 8
c + d = 12
a + c = 13
b + d = 7
= 8x12x13x7 = 8736
= 38720 / 8736 = 4.432 **Pearson χ^2 = 4.43**

However, since we have **2 cells with expected frequencies < 5**, and we have a 2 x 2 table, we could run a Yates' Correction on this same data to obtain the "true" χ^2 value.
Yates' Formula:

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

$$= \frac{20(|6 - 50| - 20/2)^2}{(8)(12)(13)(7)}$$

= 44 - 10 = 34^2 = 1156 x 20 = 23120
= 8736
= 23120 / 8736 = **Yates' Correction for Continuity χ^2 = 2.65**

If we had ignored our assumptions and gone with $\chi^2 = 4.43(1)$, we would reject the H_0, based on the chi square table, because at the .05 level, this value is in the critical region. That is, a chi square distribution with df = 1 only has 5% (i.e., .05) of values that are larger than 3.84, which is what we have obtained. However, when we ran the Yates' Correction, due to a broken assumptions, its value of 2.65(1) fell outside of the critical region of 3.84 for df = 1 and we, therefore, fail to reject the H_0.
So, you can see the conservative nature of the Yates' Correction in this instance. It is at the discretion of the researcher, regarding their context and questions posed, to decide which method to use.

McNemar's Test for Dependent Samples/Repeated Measures:

When the proportions are related (paired), the chi-square test is no longer valid since the observations in the contingency table are not independent of one another.
McNemar is a version of χ^2 for change in repeated measures (pretest-posttest or before and after situations) research. McNemar tests the direction of change and has degrees of freedom equal to 1. Because of these properties, we can use the critical values from the chi-square distribution table, as well as the same effect sizes, such as phi, for 1 df.

Assumptions:

1. Dependent measures; same subjects.
2. Two dichotomous measures.
3. $n \geq 10$.

H_0: Equal changes in both directions
H_1: Changes are not equal in both directions

An Example of Non-Symmetrical Rows and Columns Set-Up:

A sample of 100 people are asked if they receive the USA Today newspaper at their home, where 30 do receive the newspaper and 70 do not receive it (i.e., the Before). After the sample is subjected to heavy phone calling extolling the virtues of having USA Today in the home, they are checked again, where this time 70 receive the newspaper and 30 do not receive it (i.e., the After). Is there a statistically significant difference between the pre- (before) and post-test (after) conditions? Alpha = .01.

Looking at the contingency table below, we see that of the 30 people who received the newspaper before the phone calling, 20 still retained the paper after the calling, with 10 people who did not. Of the 70 people who did not receive the newspaper before the phone calling, 50 received the paper after the calling, with 20 who did not.

		Post or After		
		No Paper (-)	Have Paper (+)	
Pre or Before	Have Paper (+)	change 10 A	no change B 20	30 = A+B
	No Paper (-)	no change 20 C	change D 50	70 = C+D
		30 = A+C	B+D = 70	

Note: A and D demonstrate change
B and C demonstrate no change

McNemar's Test = $\frac{(A-D)^2}{A+D}$

$$= \frac{(10-50)^2}{10+50} = \frac{(40)^2}{60} = \frac{1600}{60} = \mathbf{26.67}$$

$x^2_{.01,1} = 6.64$

Phi = 26.67 / 100 = .2667
SQRT (.2667) = **.516**

Since we stated a χ^2 (crit) α =.01, the results are statistically significant and we reject the H_0 because the obtained chi-square value (i.e., 26.67) exceeded the critical value of 6.64. That is, the difference in score change was statistically significant between the before and after conditions, and the phi effect size (.516) for 1 df indicated that there was a large relationship between score change. Changes were not equal in both directions with the phone calling initiative allowing for a greater proportional change in the direction of people obtaining more newspapers.

Yates Correction for the McNemar Test:

An Example of Non-Symmetrical Rows and Columns Set-Up:

$$\chi^2 = \frac{([A-D]-1)^2}{A+D}$$

Pre or Before	**Post or After:** No Paper (-)	**Post or After:** Have Paper (+)	
Have Paper (+)	change 10 A	no change B 20	30 = A+B
No Paper (-)	no change 20 C	change D 50	70 = C+D
	30 = A+C	B+D = 70	

= |10 – 50| = 40 – 1 = 39^2 = 1521
= 10 + 50 = 60
= 1521 / 60 = **25.35** or $x^2_{.01,1} = 6.64$

Since we stated a χ^2 (crit) α =.01, the results are statistically significant and we reject the H_0 because the obtained chi-square value (i.e., 25.35) exceeded the critical value of 6.64. That is, the difference in score change was statistically significant between the before and after conditions.

McNemar's Test for Dependent Samples/Repeated Measures:

An Example of Symmetrical Rows and Columns Set-Up:

		Treatment 1 Yes	Treatment 1 No
Treatment 2	Yes	17	5
	No	15	3

McNemar's Test = $\frac{(B-C)^2}{B+C}$

$= \frac{(5-15)^2}{5+15} = \frac{(10)^2}{20} = \frac{100}{20} = \mathbf{5.00}$

$x^2_{.05,1} = 3.84$

Phi = 5.00 / 40 = .125
SQRT (.125) = **.354**

Since we stated a χ^2 (crit) α =.05, the results are statistically significant and we reject the H_0 because the obtained chi-square value (i.e., 5.00) exceeded the critical value of 3.84. That is, the difference in Yes/No score change was statistically significant between the Treatment 1 and Treatment 2 conditions, and the phi effect size (.354) for 1 df indicated that there was a moderate relationship between Yes/No change, where changes were not equal in both directions.

Yates Correction for the McNemar Test:

An Example of Symmetrical Rows and Columns Set-Up:

		Treatment 1 Yes	Treatment 1 No
Treatment 2	Yes	17	5
	No	15	3

$$\chi^2 = \frac{([B-C]-1)^2}{B+C}$$

= |5 – 15| = 10 – 1 = 9^2 = 81
= 5 + 15 = 20
= 81 / 20 = **4.05** or $x^2_{.05,1} = 3.84$

Phi = 4.05 / 40 = .101
SQRT (.101) = **.318**

Since we stated a χ^2 (crit) α =.05, the results are statistically significant and we reject the H_0 because the obtained chi-square value (i.e., 4.05) exceeded the critical value of 3.84. That is, the difference in Yes/No score change was statistically significant between the Treatment 1 and Treatment 2 conditions, and the phi effect size (.318) for 1 df indicated that there was a moderate relationship between Yes/No change, where changes were not equal in both directions.

Fisher's Exact Test of Significance

- Uses exact probabilities, so no minimum cell sizes are required.
- Often used when expected frequencies are **< 1** in a 2 x 2 chi square table.
- Also used when the number of observations is ≤ **20** in a 2 x 2 chi square table.

$$\textbf{Fisher's Exact: } p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{N!\,a!b!c!d!}$$

	Educational Researcher	**Education**
Reading	5 (a)	0 (b)
Math	1 (c)	4 (d)

! = factorial
- A factorial is the product of all of the whole numbers, except zero, that are less than or equal to that number.

5! = 5 x 4 x 3 x 2 x 1 = 120 or 3! = 3 x 2 x 1 = 6 or 0! = 1

Example: Let X be an educational journal, either *Educational Researcher* or *Education*, and let Y be the number of articles in each journal pertaining to Reading or Math. So, we know that *Educational Researcher* has 5 articles about Reading and 1 concerning Math, while *Education* had 0 on Reading and 4 related to Math.

a+b = 5!
c+d = 5!
c+a = 6!
b+d = 4!
= 5!(120) x 5!(120) x 6!(720) x 4!(24) = 248832000
= 10!(3628800) x 5!(120) x 1! x 1! x 4!(24) = 10450944000
= 248832000 / 10450944000
= **.023**
Since the p-value for the Fisher's Exact Test = .023, we would conclude that the distribution in the observed 2 x 2 table at the .05 level is statistically significant and different from chance. Thus, we reject H_0 and can say that there is a statistically significant association between the journal and type of article.

Below in the table, are some formulas for calculating indices for effect size and measure of agreement for **2 x 2 tables**. SPSS provides the following as choices from its menu: Yule's Q (i.e., this is also known as Gamma for ordinal measures), Phi (i.e., for nominal measures), Contingency Coefficient (i.e., for nominal measures), Somers' D_{xy} (i.e., where x is considered the dependent measure and for ordinal measures), Somers' D_{yx} (i.e., where y is considered the dependent measure and for ordinal measures), and Kappa (i.e., a measure of agreement).

TABLE 1
Indices of Association for 2 × 2 Data

Name of Coefficient	Symbol	Formula
Yule's Colligation	YC	$(\alpha^{1/2} - 1)/(\alpha^{1/2} + 1)$
Yule's Q[a]	Q	$(\alpha - 1)/(\alpha + 1)$
Pearson's Q_4	Q_4	$\sin[(\pi/2)\{1/(1 + 2bc/(ad - bc))\}]$
Pearson's Q_5	Q_5	$\sin[(\pi/2)(1/(1 + k^2)^{1/2}]$ where $k = 4abcd/(ad - bc)^2(a + d)(b + c)$
Chambers' r_e	r_e	$[(\alpha + 1)/(\alpha - 1)] - [(2\alpha\log_e\alpha)/(\alpha - 1)^2]$
Cosine approximation to the tetrachoric	r_{tc}	$\cos[\pi/(\alpha^{1/2} + 1)] = \sin[\pi(\alpha^{1/2} - 1)/2(\alpha^{1/2} + 1)]$
.74 approximation to the tetrachoric	$r_{t.74}$	$(\alpha^{.74} + 1)/(\alpha^{.74} + 1)$
Phi	ϕ	$(\chi^2)^{1/2} = ad - bc/[(a + b)(c + d)(a + c)(b + d)]^{1/2}$
Phi/phi max	ϕ/ϕ max	if $(c + d) > (b + d)$; ϕ max = $[(b + d)(a + b)/(c + d)(a + c)]^{1/2}$ if $(c + d) < (b + d)$; ϕ max = $[(c + d)(a + c)/(b + d)(a + b)]^{1/2}$
Contingency coefficient	C	$[\chi^2/(\chi^2 + 1)]^{1/2} = \left[\frac{1 + (a + b)(c + d)(a + c)(b + d)}{(a + b)(c + d)(a + c)(b + d)}\right]^{1/2}$
Kappa	κ	$[(a + d) - (a + d)(a + c) - (c + d)(b + a)]/[1 - (a + b)(a + c) - (c + d)(b + a)]$
Somers' d_{xy}	Sd_{xy}	$(ad - bc)/(ad + bc + ac + bd)$
Somers' d_{yx}	Sd_{yx}	$(ad - bc)/(ad + bc + ab + cd)$
Agreement Index	G	$2(a + d) - 1$

Note: $\alpha = ad/bc$
[a] for 2 × 2 table, Q = Goodman & Kriskal's γ.