Chi-Square (χ^2) Goodness-of-Fit

The Z-test, t-test, and Pearson's r all assume that at least one of the variables (usually the dependent variable) is measured on the interval or ratio scale. When variables of interest are nominal or categorical, these statistical tests could be inappropriate and produce misleading information. A chi-square statistic, however, provide a more appropriate assessment for such data. Two types of chi-square tests are covered in this course: goodness-of-fit and the contingency table chi-square (or test of association). Goodness-of-fit tests are designed to assess the distribution of one variable, and contingency table chi-square tests are appropriate for assessing the association between two categorical variables.

Goodness-of-Fit (one nominal or categorical variable present)

The goodness-of-fit chi-square test is designed to determine whether the frequencies observed for categories of one variable follow a pattern that departs from what is expected.

1. Example 1: Sex Distribution Within a Class

With the goodness-of-fit test one is usually interested in determining whether a given distribution of data follows an expected pattern. For example, suppose one wishes to know whether the sex distribution of students in a class is random according to the distribution found within the school. Students in the school are about 50% female and 50% male. If students are assigned at random to classes one should therefore expect about a 50:50 split in females and males within a given classroom. One classroom, however, appears to have a student sex distribution that may not be the result of random assignment, or so thought the classroom teacher who seems to have a disproportionate number of troublesome boys.

2. Hypotheses

The null hypothesis states that the distribution of students by sex within a class fits the larger distribution found within the school, i.e.

H_0: males and females are equally distributed,

or

H_0: $f_1 = f_2$

or

H_0: $\text{distribution}_{\text{pop}} = \text{distribution}_{\text{theory}}$

and the alternative hypothesis is:

H_1: not H_0; student sex not equally distributed,

or

H_1: the frequencies f_1 and f_2 are not equal

or

H_1: $\text{distribution}_{\text{pop}} \neq \text{distribution}_{\text{theory}}$

3. Observed Data for Classroom Sex Distribution

The observed frequencies of students by sex in this class are:

Table 1
Observed Counts

	Observed
Female	6
Male	16

Total = 22

Does this distribution seem likely if indeed students are randomly assigned to classes? Given the sex distribution of students within the school (where ½ are male and ½ are female), one should expect half to be females and half to be males.

Table 2
Counts and Proportions

	Observed	Proportion Expected	Expected Frequency (Total Frequency × Proportion)
Female	6	.5	22 × .5 = 11
Male	16	.5	22 × .5 = 11

Total = 22

4. Calculating χ^2 (chi-square)

The chi-square goodness-of-fit statistic to test H_0 can be calculated using the following formula:

$$\chi^2 = \sum \frac{(O_j - E_j)^2}{E_j} = n.\sum \frac{(p_j - \pi_j)^2}{\pi_j}$$

The formula on the far right is presented by some authors and is based upon proportions. The formula on the left is presented in most other textbooks and is based upon frequencies. The formula on the left is the one to be used for the remainder of this section.

The chi-square goodness-of-fit formula can be explained as follows:

(1) j = the unique cells or categories in the table of frequencies;
(2) O = the observed frequency in cell j;
(3) E = the expected frequency in cell j;
(4) $\sum$ = a summation sign—add all squared terms once division has occurred;

The expected frequencies, E_j, are determined by theory. In the example above, one may expect students to be distributed equally by sex, so 22/2 = 11 for each of the two possible cells.

The value of χ^2 is obtained as follows:

$$\chi^2 = \frac{(6-11)^2}{11} + \frac{(16-11)^2}{11}$$

$$= \frac{25}{11} + \frac{25}{11}$$

= 2.273 + 2.273

= 4.546

5. Yates' Correction (supplemental information)

Some software may present a corrected χ^2 value that adjusts for small samples or small cells sizes [count < 5 per cell]. This adjustment is known as the correction for continuity, or Yates' correction and the formula is simply to subtract -.5 from each O-E deviation before squaring:

$$\chi^2 = \sum \frac{(|O_j - E_j| - .5)^2}{E_j}$$

Note that one must use the absolute value of $|O_j\text{-}E_j|$ for the formula to work properly. In most cases this correction is not needed; sometimes this correction may over-correct and present values that are too small. Also, this correction is applied usually to χ^2 tests with df = 1 (2 cells for goodness-of-fit and 2x2 tables for tests of association [discussed elsewhere]).

6. Chi-square Distribution

The χ^2 distributions are (a) positively skewed, (b) have a minimum of zero, and (c) have just one parameter which is their degree of freedom (df). Figure 1 provides an example of a chi-square distribution for 6 degrees of freedom.

7. Degrees of freedom

The df for goodness-of-fit chi-squares is defined as:

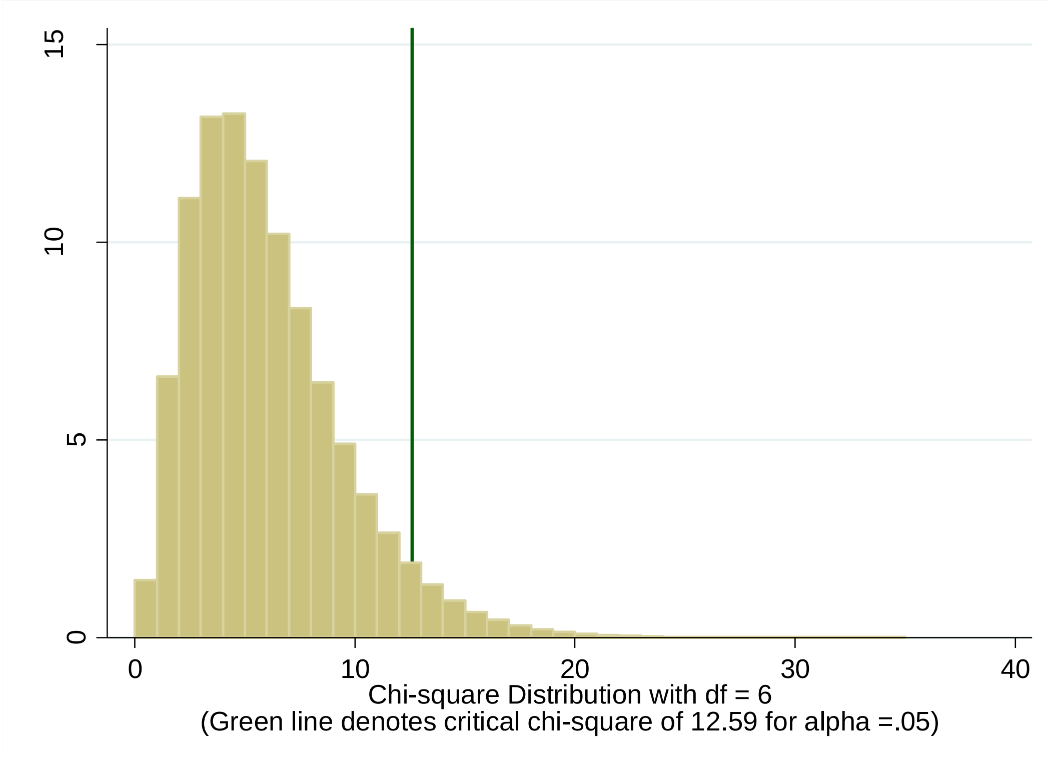
df (or ν) = J - 1

where J is the number of categories present. Since there were two categories in the example data (i.e., males and females), there is

df = 2 - 1 = 1

one df.

Figure 1
χ^2 Distribution with df = 6 (note line demarking critical value of 12.59 for α=.05)



8. Testing H_0

To statistically test the tenability of the null hypothesis, one must determine whether the calculated value of χ^2 exceeds what would be expected by chance given that H_0 is true, i.e., does the calculated χ^2 exceed the critical value of χ^2?

The critical χ^2 or ${}_{crit}\chi^2$, can be found in critical χ^2 table (see course web page).

If α = .05, the critical value for the example data is

${}_{crit}\chi^2 = 3.841$.

To test H_0, simply compare the obtained χ^2 against the critical, and if the obtained is larger, then reject H_0.

9. Decision Rule

If $\chi^2 \geq {}_{crit}\chi^2$, then reject H_0, otherwise FTR H_0.

With the current example, the decision rule is:

If 4.546 ≥ 3.841, then reject H_0, otherwise FTR H_0.

So reject the null (at alpha equal to .05) and conclude that the distribution of students by sex does not appear to be randomly distributed; i.e., there appears to be more males than females in this class and the numbers are larger than one would expect by chance alone.

10. APA Style Presentation

Goodness-of-fit results can be reported either as text or table; both approaches are illustrated below.

(a) Table

Table 3
Frequencies of Students by Sex

	Student Sex	
	Female	Male
Observed Freq.	6	16
Expected Freq. (prop.)	11 (.5)	11 (.5)

Note. χ^2 = 4.55*, df =1. Numbers in parentheses, (), are expected proportions.
Freq. = frequency and prop. = proportion.
*p < .05

Results of the goodness-of-fit test indicate that the frequencies of students by sex are not equally distributed within this class; frequencies are statistically different from what would be expected by chance. It appears that males are disproportionately over-represented in this class and females are under-represented.

(b) Text

If presenting these results within text, it could be written as follows:

The statistical results, χ^2 (1, n = 22) = 4.55, $p < .05$, indicate that the frequencies of students by sex are not equally distributed within this class; frequencies are statistically different from what would be expected by chance. It appears that males (n = 16) are disproportionately over-represented in this class and females (n = 6) are under-represented.

11. Example 2: Sex Distribution with Unequal Proportions

Instead of the school population consisting of about 50% female and 50% male, assume the school population has more males than females. Specifically, assume that females represent 33% of students and males represent 67%. If students are assigned at random to classes one should therefore expect about 2/3 of students to be male.

What impact will this have on the expected frequencies? The revised table of expected frequencies appears below.

Table 4
Sex Observed and Expected

	Observed	Proportion Expected	Expected Frequency (Total Frequency × Proportion)
Female	6	.33	22 × .33 = 7.26
Male	16	.67	22 × .67 = 14.74

Total = 22

12. Calculating χ^2 (chi-square)

Using these revised proportions and expected frequencies result in the following χ^2 value, using the same formula:

$$\chi^2 = \sum \frac{(|O_j - E_j| - .5)^2}{E_j}$$

$$\chi^2 = \frac{(6-7.26)^2}{7.26} + \frac{(16-14.74)^2}{14.74}$$

$$= \frac{1.588}{7.26} + \frac{1.588}{14.74}$$

= 0.219 + 0.108

= 0.327

13. Degrees of freedom and Critical Value

The df remain the same at

df = 2 - 1 = 1

so the critical value will also remain the same for α = .05:
one df.

${}_{crit}\chi^2 = 3.841$.

The decision rule now states the following:

If 0.327 ≥ 3.841, then reject H_0, otherwise FTR H_0.

Since 0.327 is less than the critical value of 3.841 one would fail to reject the null and therefore conclude that the sex distribution of students within the class are similar to the distribution found within the school.

14. Example 3: Birth Distribution

As a second example suppose one wishes to know whether the distribution of births throughout the year is random, with equal probabilities or frequencies during the year.

15. Hypotheses

The null hypothesis states that the distribution of births throughout the year is random and has equal frequencies, i.e.

H_0: frequency of births equally distributed throughout year,

or

H_0: $f_1 = f_2 = f_3 = ... = f_6$

or

H_0: distribution$_{pop}$ = distribution$_{theory}$

and the alternative hypothesis is:

H_1: not H_0; births not equally distributed throughout year,

or

H_1: the frequencies f_1, f_2, ..., f_6 are not all equal

or

H_1: distribution$_{pop}$ $\neq$ distribution$_{theory}$

16. Observed Data for Birth by Month Distribution

Assume the researcher obtained information about frequencies of birth from the local hospital. The following frequencies were observed:

Table 5
Observed Births

	Jan.-Feb.	March-April	May-June	July-August	Sept.-October	Nov.-Dec.
Observed	71	78	83	94	112	114

Total births = 552

Does this distribution seem random if indeed births are equally likely throughout the year; that is, does the null hypothesis of equality of births throughout the year seem tenable?

Table 6
Observed and Expected Births

	Jan.-Feb.	March-April	May-June	July-August	Sept.-October	Nov.-Dec.
Observed	71	78	83	94	112	114
Proportion Expected	1/6 (.16667)	1/6 (.16667)	1/6 (.16667)	1/6 (.16667)	1/6 (.16667)	1/6 (.16667)
Expected Frequency (Total Frequency × Proportion)	552 × .16667 = 92	552 × .16667 = 92	552 × .16667 = 92	552 × .16667 = 92	552 × .16667 = 92	552 × .16667 = 92

Total births = 552; expected = 552/6 = 92

17. Calculating χ^2 (chi-square)

As noted above, the chi-square goodness-of-fit statistic to test H_0 can be calculated using the following formula:

$$\chi^2 = \sum \frac{(O_j - E_j)^2}{E_j}$$

The expected frequencies, E_j, are determined by theory. In the example above, it was expected that the frequencies of births would be equally distributed across the year. Since 552 births were observed during the year, this means that 552/6 = 92 births are expected every two months. If, for example, one wanted to find the expected frequency for every month, the expected frequency would be 552/12 = 46.

The value of χ^2 is obtained as follows:

$$\chi^2 = \frac{(71-92)^2}{92} + \frac{(78-92)^2}{92} + \frac{(83-92)^2}{92} + \frac{(94-92)^2}{92} + \frac{(112-92)^2}{92} + \frac{(114-92)^2}{92}$$

$$= \frac{441}{92} + \frac{196}{92} + \frac{81}{92} + \frac{4}{92} + \frac{400}{92} + \frac{484}{92}$$

$$= 4.793 + 2.130 + .880 + .043 + 4.348 + 5.261$$

$$= 17.455$$

18. Degrees of freedom

Since there were six categories in the example data, there are

df = 6 - 1 = 5

five df.

19. Testing H_0

The critical χ^2 or ${}_{crit}\chi^2$, can be found in Table 1 above. If $\alpha = .05$, the critical value for the example data is

${}_{crit}\chi^2 = 11.07$.

To test H_0, compare the obtained χ^2 against the critical, and if the obtained is larger, then reject H_0.

20. Decision Rule

If $\chi^2 \geq {}_{crit}\chi^2$, then reject H_0, otherwise FTR H_0.

With the current example, the decision rule is:

If 17.455 ≥ 11.07, then reject H_0, otherwise FTR H_0.

So reject the null (at alpha equal to .05) and conclude that the distribution of births does not appear to be randomly distributed; i.e., the distribution of births seems to be higher during the fall and earlier winter months than during other times of the year.

21. APA Style Presentation

(a) Table

Table 7
Frequencies of Births by Months

	Jan.-Feb.	Mar.-Apr.	May-June	July-Aug.	Sept.-Oct.	Nov.-Dec.
Observed Freq.	71	78	83	94	112	114
Expected Freq. (prop.)	92 (.167)	92 (.167)	92 (.167)	92 (.167)	92 (.167)	92 (.167)

Note. χ^2 = 17.456*, df =5. Numbers in parentheses, (), are expected proportions. Freq. = frequency and prop. = proportion.
*p < .05

The goodness-of-fit results indicate statistical differences in birth frequencies throughout the year. Based upon the observed frequencies it appears that the birth rate is highest for the months of September to December, and lowest for the spring and summer months.

(b) Text

If presenting these results within text, it could be written as follows:

The goodness-of-fit results, χ^2 (5, n = 552) = 17.455, *Section* , indicate statistical differences in birth frequencies throughout the year. Based upon the observed frequencies it appears that the birth rate is highest for the months of September to December, and lowest for the spring and summer months.

22. Exercises

1. Horse-racing fans often maintain that in a race around a circular track significant advantages accrue to the horses in certain post positions. Any horse's post position is his assigned post in the starting line-up. Position 1 is closest to the rail on the inside of the track; and position 8 is on the outside, farthest from the rail in an 8-horse race. Test whether post position is related to race results. (This example taken from S. Siegel, 1956, Nonparametric statistics for the behavioral sciences, McGraw-Hill.)

Listed below are the observed frequencies of 1st place finishers for each post position during a regular month at the tracks.

Post Position

	1	2	3	4	5	6	7	8
No. of wins	29	19	18	25	17	10	15	11

Total wins = 144.

(a) What are the null and alternative hypotheses?
(b) What are the expected frequencies if post position is unrelated to winning?
(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .05 level?
(d) What are the results of this test?

2. The director of athletics at the local high school wonders if the sports program is getting a proportional amount of support from each of the four classes represented in the high school. If there are roughly equal numbers of students in each of the classes, what does the following breakdown of attendance figures from a random sample of students in attendance at a recent basketball game suggest? (This example taken from J. F. Healey, 1993, Statistics: A tool for social research, 3rd ed., Wadsworth.)

Class	Frequency
Freshmen	200
Sophomores	150
Juniors	120
Seniors	110

Total attendance = 580.

(a) What are the null and alternative hypotheses?
(b) What are the expected frequencies for attendance by class?
(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .01 level?
(d) What are the results of this test?

3. Using the same data as above, assume that class rank is not equally distributed within the high school. Instead, about 30% of students are freshmen, 28% are sophomores, 22% are juniors, and 20% are seniors. Using these percentages to estimate expected values for attendance at the basketball game, is there any evidence that attendance differs from what one would expect given the class rank distribution of students enrolled in the school?

4. Suppose someone has a hypothesis that the "Transylvania effect" of the full moon is related to incidence of drug overdose. A search of medical files at a hospital yielded 1182 drug overdose cases which included the date. The full-moon phase was based on the actual dates of the full moon plus or minus 2 days, yielding 75 full-moon days and 381 non-full-moon days (total number of days was 456 for the period investigated). The observed frequency of drug overdose was 196 during the full-moon days and 986 during the non-full-moon days.

(a) What are the null and alternative hypotheses?
(b) What are the expected frequencies?
(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .05 level?
(d) What are the results of this test?

Answers are provided below.

23. Exercise Answers

1. Wins by Pole Position

(a) What are the null and alternative hypotheses?

Null: Frequency of wins are equal across pole positions.
Alternative: Frequency of wins are unequal across pole positions.

(b) What are the expected frequencies if post position is unrelated to winning?

144/8 = 18, or .125 per pole

(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .05 level?

$\chi^2 = 16.33$, df = 7, p = .0223

(d) What are the results of this test?

Reject Ho and conclude that winning varies by pole position. It seems there are more wins than one would expect for position 1 and 4 (residuals or deviations are greatest here) and fewer wins that would be expected for positions 6 and 8 (again, large deviations here).

2. Athletic Attendance

(a) What are the null and alternative hypotheses?

Null: Student attendance equal by class rank
Alternative: Student attendance unequal by class rank

(b) What are the expected frequencies for attendance by class?

580/4 = 145, or .25 per class rank

(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .01 level?

$\chi^2 = 33.79$, df = 3, p = .0001

(d) What are the results of this test?

Reject Ho and conclude that attendance differs by class rank. Examination of deviation scores shows that freshmen attend at a higher rate than would be expected and seniors attend at a lower rate than would be expected.

3. Athletic Attendance with Proportional Attendance

(a) What are the null and alternative hypotheses?

Null: Student attendance proportional to school enrollment.
Alternative: Student attendance disproportional to school enrollment.

(b) What are the expected frequencies for attendance by class?

Freshmen = 174 or .3 of attendance
Sophomores = 162.4 or .28 of attendance

Juniors = 127.6 or .22 of attendance
Seniors = 116 or .20 of attendance

(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .01 level?

$\chi^2 = 5.59$, df = 3, p = .1334

(d) What are the results of this test?

Fail to reject Ho and conclude that attendance appears to follow a distribution similar to general enrollment in the school.

4. Drug overdoses and Transylvania Effect

Suppose someone has a hypothesis that the "Transylvania effect" of the full moon is related to incidence of drug overdose. A search of medical files at a hospital yielded 1182 drug overdose cases which included the date. The full-moon phase was based on the actual dates of the full moon plus or minus 2 days, yielding 75 full-moon days and 381 non-full-moon days (total number of days was 456 for the period investigated). The observed frequency of drug overdose was 196 during the full-moon days and 986 during the non-full-moon days.

(a) What are the null and alternative hypotheses?

Null: Overdoses are proportion to number of days exposed.
Alternative: Overdoses are disproportion to number of days exposed.

(b) What are the expected frequencies?

Total number of overdoses = 196+986 = 1182

Total days = 75 + 381 = 456
Days of full moon = 75/456 = .1645
Days of non-full moon = 381/456 = .8355

Expected frequencies for overdoses would be total overdoses × proportions above, so
Overdoses on full moon days = .1645 × 1182 = 194.439
Overdoses on non-full moon days = .8355 × 1182 = 987.561

(c) What is the obtained and critical chi-square statistics and df if alpha is set at the .05 level?

$\chi^2 = 0.01$, df = 1, p = .92

(d) What are the results of this test?

Results show that drug overdoses occur at the expected frequency given the number of days considered, so it seems there is no full-moon effect on drug overdoses.