Chi-Square Test for Goodness of Fit

(after *Applied Statistics* by Hinkle/Wiersma/Jurs)

Scientists will often use the Chi-square (χ^2) test to determine the goodness of fit between theoretical and experimental data. In this test, we compare **observed values** with **theoretical or expected values.** Observed values are those that the researcher obtains empirically through direct observation; theoretical or expected values are developed on the basis of some hypothesis. For example, in 200 flips of a coin, one would expect 100 heads and 100 tails. But what if 92 heads and 108 tails are observed? Would we reject the hypothesis that the coin is fair? Or would we attribute the difference between observed and expected frequencies to random fluctuation?

Consider another example. Suppose we hypothesize that we have an unbiased six-sided die. To test this hypothesis, we roll the die 300 times and observe the frequency of occurrence of each of the faces. Because we hypothesized that the die is unbiased, we expect that the number on each face will occur 50 times. However, suppose we observe frequencies of occurrence as follows:

Face value	*Occurrence*
1	42
2	55
3	38
4	57
5	64
6	44

Again, what would we conclude? Is the die biased, or do we attribute the difference to random fluctuation?

Consider a third example. The president of a major university hypothesizes that at least 90 percent of the teaching and research faculty will favor a new university policy on consulting with private and public agencies within the state. Thus, for a random sample of 200 faculty members, the president would *expect* 0.90 x 200 = 180 to favor the new policy and 0.10 x 200 = 20 to oppose it. Suppose, however, for this sample, 168 faculty members favor the new policy and 32 oppose it. Is the difference between observed and expected frequencies sufficient to reject the president's hypothesis that 90 percent would favor the policy? Or would the differences be attributed to chance fluctuation?

Lastly, consider an experimental result where, for given independent values of "X," the following theoretical (expected) and experimental (observed) dependent values of Y were found:

X	$Y_{theoretical}$	$Y_{experimental}$
3.45	11.90	11.37
4.12	16.97	17.02
4.73	22.37	23.78
5.23	27.35	26.13
6.01	36.12	35.96
6.82	46.51	45.22
7.26	52.71	53.10

In each of these examples, the test statistic for comparing observed and expected frequencies is χ^2, defined as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

where

O = observed value
E = expected value
k = number of categories, groupings, or possible outcomes

The calculations of χ^2 for each of the three examples, using the above formula, are found in Tables 1-4.

TABLE 1. Calculation of χ^2 for the Coin-Toss Example

Face	**O**	**E**	**O-E**	**(O-E)^2**	**(O-E)^2/E**
Heads	92	100	-8	64	.64
Tails	108	100	+8	64	.64
Totals	200	200	0	--	1.28 = χ^2

TABLE 2. Calculation of χ^2 for the Die Example

Face Value	**O**	**E**	**O-E**	**(O-E)^2**	**(O-E)^2/E**
1	42	50	-8	64	1.28
2	55	50	5	25	.50
3	38	50	-12	144	2.88
4	57	50	7	49	.98
5	64	50	14	196	3.92
6	44	50	-6	36	.72
Totals	300	300	0	--	10.28= χ^2

TABLE 3. Calculation of χ^2 for the Consulting-Policy Example

Disposition	**O**	**E**	**O-E**	**(O-E)^2**	**(O-E)^2/E**
Favor	168	180	-12	144	.80
Oppose	32	20	+12	144	7.20
Totals	200	200	0	--	8.00= χ^2

TABLE 4. Calculation of χ^2 for the Experiment Example

X	**O**	**E**	**O-E**	**(O-E)^2**	**(O-E)^2/E**
3.45	11.37	11.90	-.53	.28	.02
4.12	17.02	16.97	.05	.02	.00
4.73	23.78	22.37	1.41	1.99	.09
5.23	26.13	27.35	-1.22	1.49	.05
6.01	35.96	36.12	-.16	.03	.00
6.82	45.22	46.51	-1.29	1.66	.04
7.26	53.10	52.71	.39	.15	.00
Totals	212.58	213.93	-1.35	5.62	0.20 = χ^2

There is a family of χ^2 distributions, each a function of the degrees of freedom associated with the number of categories in the sample data. Only a single degree of freedom (df) value is required to identify the specific χ^2 distribution. Notice that all values of χ^2 are positive, ranging from zero to infinity.

Consider the degrees of freedom for each of the above examples. In the coin example, note that the *expected* frequencies in each of the two categories (heads or tails) are *not* independent. To obtain the expected frequency of tails (100), we need only to subtract the expected frequency of heads (100) from the total frequency (200), or 200 - 100 = 100. Similarly, for the example of the new consulting policy, the expected number of faculty members who oppose it (20) can be found by subtracting the expected number who support it (180) from the total number in the sample (200), or 200 – 180 = 20. Thus, given the expected frequency in one of the categories, the expected frequency in the other is readily determined. In other words, only the expected frequency in one of the two categories is free to vary; that is, there is only 1 degree of freedom associated with these examples.

For the die example, there are six possible categories of outcomes: the occurrence of the six faces. Under the assumption that the die is fair, we would *expect* that the frequency of occurrence of each of the six faces of the die would be 50. Note again that the expected frequencies in each of these categories are *not* independent. Once the expected frequency for five of the categories is known, the expected frequency of the sixth category is uniquely determined, since the total frequency equals 300. Thus, only the expected frequencies in five of the six categories are free to vary; there are only 5 degrees of freedom associated with this example.

In the last example dealing with the data from the experiment, the cross-tabulation table has 2 columns and 7 rows. The degrees of freedom for such tables is (# rows – 1)(# columns – 1) = (6)(1) = 6

The Critical Values for the χ^2 Distribution

The use of the χ^2 distribution in hypothesis testing is analogous to the use of the *t* and F distributions. A null hypothesis is stated, a test statistic is computed, the observed value of the test statistic is compared to the critical value, and a decision is made whether or not to reject the null hypothesis. For the coin example, the null hypothesis is that the frequency of heads is equal to the frequency of tails. For the die example, the null hypothesis is that the frequency of occurrence of each of the six faces is the same. In general, it is not a requirement for the categories to have equal expected frequencies. For instance, in the example of the new consulting policy, the null hypothesis is that 90 percent of the faculty will support the new policy and 10 percent will not.

The critical values of χ^2 for 1 through 30 degrees of freedom are found in Table 5. Three different percentile points in each distribution are given – $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$. (e.g., chances of 10%, 5%, and 1% respectively of rejecting the null hypothesis when it should be retained). For the coin and consulting-policy examples, the critical values of χ^2 for 1 degree of freedom, with $\alpha = .05$ and $\alpha = .01$, are 3.841 and 6.635, respectively. For the die example, the corresponding critical values of χ^2 for 5 degrees of freedom are 11.070 and 15.086. Although Table 5 is sufficient for many research settings in the sciences, there are some situations in which the degrees of freedom associated with a χ^2 test are greater than 30. These situations are not addressed here.

Now that we have seen the table of critical values for the χ^2 distribution, we can complete the examples. For the coin example, the null hypothesis is that the frequency of heads equals the frequency of tails. As we mentioned, because there are only two categories, once the expected value of the first category is determined, the second is uniquely determined. Thus, there is only 1 degree of freedom associated with this example. Assuming that the $\alpha = .05$ level of significance is used in testing this null hypothesis, the critical value of χ^2 (χ^2_{CV}) is 3.841 (see Table 5). Notice that, in Table 1, the calculated value of χ^2 is 1.28. Because the calculated value does *not* exceed the critical value, the null hypothesis (the coin is fair) is *not* rejected; the differences between observed and expected frequencies are attributable to chance fluctuation. That is, when the calculated value of χ^2 exceeds the critical value, the data support the belief that a significant difference exists between expected and actual values.

For the example of the new consulting policy, the null hypothesis is that 90 percent of the faculty would support it and 10 percent would not. Again, because there are only two categories, there is 1 degree of freedom associated with the test of this hypothesis. Thus, assuming $\alpha = .05$, the χ^2_{CV} is 3.841. From Table 3, we see that the calculated value of χ^2 is 8.00; therefore, the null hypothesis (that there is no difference) is rejected. The conclusion is that the percentage of faculty supporting the new consulting policy is not 90.

For the die example, the null hypothesis is that the frequency of occurrence of each of the six faces is the same. With six categories, there are 5 degrees of freedom associated with the test of this hypothesis; the χ^2_{CV} for $\alpha = .05$ is 11.070. Using the data from Table 2, $\chi^2 = 10.28$. Because this calculated value is less than the critical value, the null hypothesis is retained. The conclusion is that the differences between the observed and expected frequencies in each of the six categories are attributable to chance fluctuation.

For the example with the experiment, the null hypothesis is that there is no difference between theoretical and experimental results. With seven data pairs there are 6 degrees of freedom associated with the test of this hypothesis; the χ^2_{CV} for $\alpha = .05$ is 12.592. Because this calculated value of χ^2 (0.20) is less than the critical value (12.592), the null hypothesis is retained. The conclusion is that the differences between the observed and expected values in each of the seven data pairs are attributable to chance fluctuation. The experimental results are therefore consistent with the theoretical results to with a 95% chance of probability.

TABLE 5. CRITICAL χ^2 VALUES FOR UP TO 30 DEGREES OF FREEDOM

df	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566
21	29.615	32.671	38.932
22	30.813	33.924	40.289
23	32.007	35.172	41.638
24	33.196	36.415	42.980
25	34.382	37.652	44.314
26	35.563	38.885	45.642
27	36.741	40.113	46.963
28	37.916	41.337	43.278
29	39.087	42.557	49.558
30	40.256	43.773	50.892