

Chapter 13 Section 1

Example 2

The chi-square distributions are a family of distributions that take only positive values and are skewed to the right. A specific chi-square distribution is specified by one parameter, called the degrees of freedom.

The figure shows the density curves for three members of the χ^2 family of distributions. As the degrees of freedom increase, the density curves become less skewed and larger values become more probable.

Table E in the back of the book gives critical values for chi-square distributions. You can use Table E if software does not give you P-values for a χ^2 test.

The chi-square density curves have the following properties:

1. The total area under a chi-square curve is equal to 1.
2. Each chi-square curve (except when $df = 1$) begins at 0 on the horizontal axis, increases to a peak, and then approaches the horizontal axis asymptotically from above.
3. Each chi-square curve is skewed to the right. As the number of degrees of freedom increase, the curve becomes more and more symmetrical and looks more like a normal curve.

GOODNESS OF FIT TEST

A **goodness of fit test** is used to help determine whether a population has a certain hypothesized distribution, expressed as proportions of population members falling into various outcome categories. Suppose that the hypothesized distribution has n outcome categories. To test the hypothesis

H_0 : the actual population proportions are *equal to* the hypothesized proportions

first calculate the chi-square test statistic

$$X^2 = \sum (O - E)^2 / E$$

Then X^2 has approximately a χ^2 distribution with $(n - 1)$ degrees of freedom.

For a test of H_0 against the alternative hypothesis

H_a : the actual population proportions are *different from* the hypothesized proportions

the P-value is $P(\chi^2 \geq X^2)$.

We use the χ^2 density curve with n-1 degrees of freedom to calculate the P-value in a goodness of fit test. The following box summarizes the details.

EXAMPLE 13.2 RED-EYED FRUIT FLIES

Biologists wish to mate two fruit flies having genetic makeup RrCc, indicating that it has one dominant gene (R) and one recessive gene (r) for eye color, along with one dominant (C) and one recessive (c) gene for wing type. Each offspring will receive one gene for each of the two traits from both parents. The following table, often called a Punnett square, shows the possible combinations of genes received by the offspring.

		Parent 1			
		RC	Rc	rC	rc
Parent 2	RC	RRCC(x)	RRCc(x)	RrCC(x)	RrCc(x)
	Rc	RRCc(x)	RRcc(y)	RrCc(x)	Rrcc(y)
	rC	RrCC(x)	RrCc(x)	rrCC(z)	rrCc(z)
	rc	RrCc(x)	Rrcc(y)	rrCc(z)	rrcc(w)

Any offspring receiving an R gene will have red eyes, and any offspring receiving a C gene will have straight wings. So based on this Punnett square, the biologists predict a ratio of 9 red-eyed, straight-wing (x):3 red-eyed, curly wing (y):3 white-eyed, straight (z):1 white-eyed, curly (w) offspring. In order to test their hypothesis about the distribution of offspring, the biologists mate the fruit flies. Of 200 offspring, 101 had red eyes and straight wings, 42 had red eyes and curly wings, 49 had white eyes and straight wings, and 10 had white eyes and curly wings.

$$9 : 3 : 3 : 1$$

Do these data differ significantly from what the biologists have predicted?

Step 1:

Identify the population of interest and the parameter(s) that you want to draw conclusions about.

State hypotheses in words and symbols.

The biologists are interested in the proportion of offspring that fall into each genetic

category for the population of all fruit flies that would result from crossing two parents with genetic makeup RrCc.

$$\left(9 + 3 + 3 + 1 = 16 \quad \frac{9}{16} = 0.5625 \quad \frac{3}{16} = 0.1875 \quad \frac{1}{16} = 0.0625 \right)$$

Their hypotheses are

$$H_o: \mathbf{p}_{red, straight} = 0.5625, \mathbf{p}_{red, curly} = 0.1875, \mathbf{p}_{white, straight} = 0.1875, \mathbf{p}_{white, curly} = 0.0625$$

H_a : at least one of these proportions is incorrect

Step 2: Choose the appropriate inference procedure and verify the conditions for using it.
We can use a chi-square goodness of fit test to measure the strength of the evidence against the hypothesized distribution, provided that the expected cell counts are large enough. Here are the expected counts:

Red-eyed, straight-wing: $200(0.5625) = 112.5$

Red-eyed, curly-wing: $200(0.1875) = 37.5$

White-eyed, straight-wing: $200(0.1875) = 37.5$

White-eyed, curly-wing: $200(0.0625) = 12.5$

Since, all the expected cell counts are greater than 5, we can proceed with the test.

Step 3: Carry out the inference procedure:

- The test statistic is

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(101 - 112.5)^2}{112.5} + \frac{(42 - 37.5)^2}{37.5} + \frac{(49 - 37.5)^2}{37.5} + \frac{(10 - 12.5)^2}{12.5} =$$

$$= 1.1756 + 0.54 + 3.5267 + 0.5 = 5.742$$

- For $\mathbf{df} = 4 - 1 = 3$, Table E shows that our test statistic falls between the critical values for a 0.15 and a 0.10 significance level. Technology produces the actual P-value of 0.1248.

Step 4: Interpret your results in the context of the problem.

The P-value of 0.1248 indicates that the probability of obtaining a sample of 200 fruit fly offspring in which the proportions differ from the hypothesized values by at least as much as the ones in our sample is over 12%, assuming that the null hypothesis is true. This is **not sufficient** evidence to reject the biologists' predicted distribution.

You can simplify the computations of Example 13.2 and other goodness of fit problems by using your calculator's list operations. The calculator also allows you to compute and visualize the P-value for a chi-square test.

- Clear lists L_1 /list1, L_2 /list2 and L_3 /list3.
- Enter the observed count in L_1 /list1. Calculate the expected counts separately and enter them in L_2 /list2.

- Define L_3 as $(L_1 - L_2)^2 / L_2$. Define list3 as $[(list1-list2)^2 / list2]$.

For TI 83

Use the command sum (L_3) to calculate the test statistic χ^2 .
(sum is located in MATHLIST.)

For Ti 89

Use the command sum(list3) to calculate χ^2 . (sum is located in the CATALOG).

- Find the P-value using the χ^2 cdf command (in the distributions (DISTR) menu on the TI-83 and in the CATALOG under Flash Apps on the TI-89).
- We ask for the area between $\chi^2 = 5.742$ and a very large number (1E99) and specify the degrees of freedom.

The P-value, 0.1248, indicates that $\chi^2 = 5.742$ is a possible result if H_0 is true, which does not provide strong enough evidence to reject H_0 .

- To visualize this P-value, first set your viewing window as shown. Then enter the command Shade χ^2 (5.742, 1E99, 3).
(This command is located in the DISTR/DRAW menu on the TI-83 and in the CATALOG under Flash Apps on the TI-89.) Note that the Shade χ^2 command requires a left endpoint and a right endpoint, so take a sufficiently large right endpoint to achieve four-decimal-place accuracy.

WINDOW

Xmin =0
Xmax =14
Xscl =1
Ymin=-.1
Ymax= .3
Yscl=.1
Xres=1