Today's Objective:
- Perform a chi square test for goodness of fit (GOF)
- Conduct a follow-up analysis when the results are statistically significant

Carrying Out a Chi-Square Test for Goodness of Fit

**CONDITIONS FOR PERFORMING A CHI-SQUARE TEST FOR GOODNESS OF FIT**

- Random: The data come from a well-designed random sample or random experiment.
  - 10%: When sampling without replacement, check that \( n \leq \frac{1}{10} N \).
- Large Counts: All expected counts are at least 5.

Remember:
Use observed and expected counts, not observed and expected proportions. If the expected counts are not all at least 5, you can combine categories until the Large Counts condition is satisfied. The Large Counts Condition here serves the same purpose as it did in z-tests for proportions: to ensure that the probability distribution that we use to calculate the P-value is a good model for the actual sampling distribution of the test statistic we are using.

THE CHI-SQUARE TEST FOR GOODNESS OF FIT

Suppose the conditions are met. To determine whether a categorical variable has a specified distribution in the population of interest, expressed as the proportion of individuals falling into each possible category, perform a test of

- \( H_0: \) The stated distribution of the categorical variable in the population of interest is correct/same
- \( H_a: \) The stated distribution of the categorical variable in the population of interest is not correct/different

Start by finding the expected count for each category assuming that \( H_0 \) is true. Then calculate the chi-square statistic

\[
\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}
\]

where the sum is over the \( k \) different categories. The P-value is the area to the right of \( \chi^2 \) under the density curve of the chi-square distribution with \( k - 1 \) degrees of freedom.
$\chi^2$ Goodness of Fit Test on the calculator:

1. Enter the observed values into List 1 and the expected values into List 2.
2. Then perform the test: STAT → TESTS → D: $\chi^2$GOF-Test (Observed: L1, Expected: L2, df: __).
3. CALCULATE or DRAW (it will actually draw the curve for you!).
4. WRITE down the procedure name, test statistic $\chi^2$, the degrees of freedom, and of course the $P$-value.
5. If your test is statistically significant ($P$-value < significance level), then look at the CNTRB components to determine which component made the largest contribution to the large chi-square statistic.

AP® EXAM TIP  You can use your calculator to carry out the mechanics of a significance test on the AP® exam. But there's a risk involved. If you just give the calculator answer with no work, and one or more of your values is incorrect, you will probably get no credit for the "Do" step. We recommend writing out the first few terms of the chi-square calculation followed by "...". This approach might help you earn partial credit if you enter a number incorrectly. Be sure to name the procedure ($\chi^2$GOF-Test) and to report the test statistic ($\chi^2 = 11.2$), degrees of freedom (df = 3), and $P$-value (0.011).

Section 11.1 Day 2 Big Ideas: Chi-Square Goodness of Fit Test

**Conditions:**

- Random
- 10%. Condition
- List expected counts & state that they are all at least 5.

**Chi-Square Distribution:**

- $\chi^2(df)$
- Always right-skewed
- Starts at 0

- $df = k - 1$
- categories

**Four Step Process:**

As repetitive as it might seem, always use the FOUR-STEP PROCESS to receive full credit on the AP® Exam!

**Follow-Up Analysis:** If test is significant (you reject Ho) then discuss which values were the largest contributors to the high $\chi^2$ value (called the "components").

The ___ groups that contributed the most to the $\chi^2$ statistic were _____ (___ more/less than expected) and ____ (more/less than expected).
EXAMPLE: Landline Surveys

According to the 2000 census, of all U.S. residents aged 20 and older, 19.1% are in their 20s, 21.5% are in their 30s, 21.1% are in their 40s, 15.5% are in their 50s, and 22.8% are 60 and older. The table below shows the age distribution for a sample of U.S. residents aged 20 and older. Members of the sample were chosen by randomly dialing landline telephone numbers. Do these data provide convincing evidence that the age distribution of people who answer landline telephone surveys is not the same as the age distribution of all U.S. residents?

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Obs.</th>
<th>Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–29</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–39</td>
<td>186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–49</td>
<td>224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59</td>
<td>211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60+</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1048</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We want to test the following hypotheses at the $\alpha = 0.05$ significance level:

$H_0$: The age distribution of people who answer landline surveys is the same as the distribution of all US residents.

$H_a$: The age distribution of people who answer landline surveys is not the same as the age distribution of all US residents.

If the conditions are met, we will perform a $\chi^2$ Goodness of Fit Test.

- Random: The data come from a random sample of US residents who answer landline surveys.
- 10% Condition: $1048 \leq 10\% \text{ (all US residents aged 20 or older)}$.
- Large Counts: The expected counts are $200.2, 225.3, 221.1, 162.4$, and $238.9$. All expected counts are at least 5.

Test Statistic: $\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(141-200.2)^2}{200.2} + \ldots + \frac{(286-238.9)^2}{238.9} = 48.2$

P-value: using $df = 5-1 = 4$, the P-value $\approx 0$.

We reject $H_0$ because the P-value $\approx 0$ is less than $\alpha = 0.05$. We have convincing evidence that the age distribution of people who answer landline surveys is different than the age distribution of all U.S. residents aged 20 or older.

Follow-Up Analysis: The two age groups that contributed the most to the $\chi^2$ statistic were the 20-to-29 year olds (59.2 fewer than expected) and the 50-to-59 year olds (48.6 more than expected).

$\checkmark$ see pg. 691
Biologists wish to mate pairs of fruit flies having genetic makeup RrCc, indicating that each has a 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 each parent. The following table, known as a Punnett square, shows the possible combinations of genes received by offspring. 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-winged offspring (x), 3 red-eyed curly-winged offspring (y), 3 white-eyed, straight-winged offspring (z), and 1 white-eyed, curly-winged (w) offspring. To test their hypothesis about the distribution of offspring, the biologists mate a random sample of pairs of fruit flies. Of 200 offspring, 99 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. Do these data differ significantly from what the biologists have predicted? Carry out a test at the \( \alpha = 0.05 \) significance level.

<table>
<thead>
<tr>
<th>Parent 1 passes on:</th>
<th>Parent 2 passes on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RC )</td>
<td>( RrCc )</td>
</tr>
<tr>
<td>( Rr )</td>
<td>( RrCc )</td>
</tr>
<tr>
<td>( rC )</td>
<td>( RrCc )</td>
</tr>
<tr>
<td>( rc )</td>
<td>( RrCc )</td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccc}
RC & RrCc & RrCc & RrCc \\
Rr & RrCc & RrCc & RrCc \\
rC & RrCc & RrCc & RrCc \\
rc & RrCc & RrCc & RrCc \\
\end{array}
\]

**Check Your Understanding, page 691:**
1. **State:** We want to perform a test at the \( \alpha = 0.01 \) significance level of \( H_0 \): the distribution of eye color and wing shape is the same as what the biologists predict versus \( H_a \): the distribution of eye color and wing shape is not the same as what the biologists predict. Or,

\[
H_0: p_{red-straight} = \frac{9}{16}, p_{red-curly} = \frac{3}{16}, p_{white-straight} = \frac{3}{16}, p_{white-curly} = \frac{1}{16} \quad \text{versus} \quad H_a: \text{At least two of the } p_i \text{s is incorrect.}
\]

**Plan:** We should use a chi-square test for goodness of fit if the conditions are met. Random: The data are from a random sample. 10\%: \( n = 200 \) is less than 10\% of all fruit flies. Large Counts: The expected counts in each category are all at least 5.

- red-straight: 200(\( \frac{9}{16} \)) = 112.5, red-curly: 200(\( \frac{3}{16} \)) = 37.5, white straight: 200(\( \frac{3}{16} \)) = 37.5, and
- white-curly: 200(\( \frac{1}{16} \)) = 12.5.

**Do:** The test statistic is

\[
\chi^2 = \frac{(99-112.5)^2}{112.5} + \frac{(42-37.5)^2}{37.5} + \frac{(49-37.5)^2}{37.5} + \frac{(10-12.5)^2}{12.5} = 6.1867. \text{ With df} = 4 - 1 = 3, \text{ the} \ P\text{-value is between 0.10 and 0.15. Using technology: } P\text{-value} = 0.1029. \text{ Conclude: Because the} \ P\text{-value of 0.1029 is greater than } \alpha = 0.01, \text{ we fail to reject } H_0. \text{ We do not have convincing evidence that the distribution of eye color and wing shape is different from what the biologists predict.}
\]

*No follow-up analysis needed.*