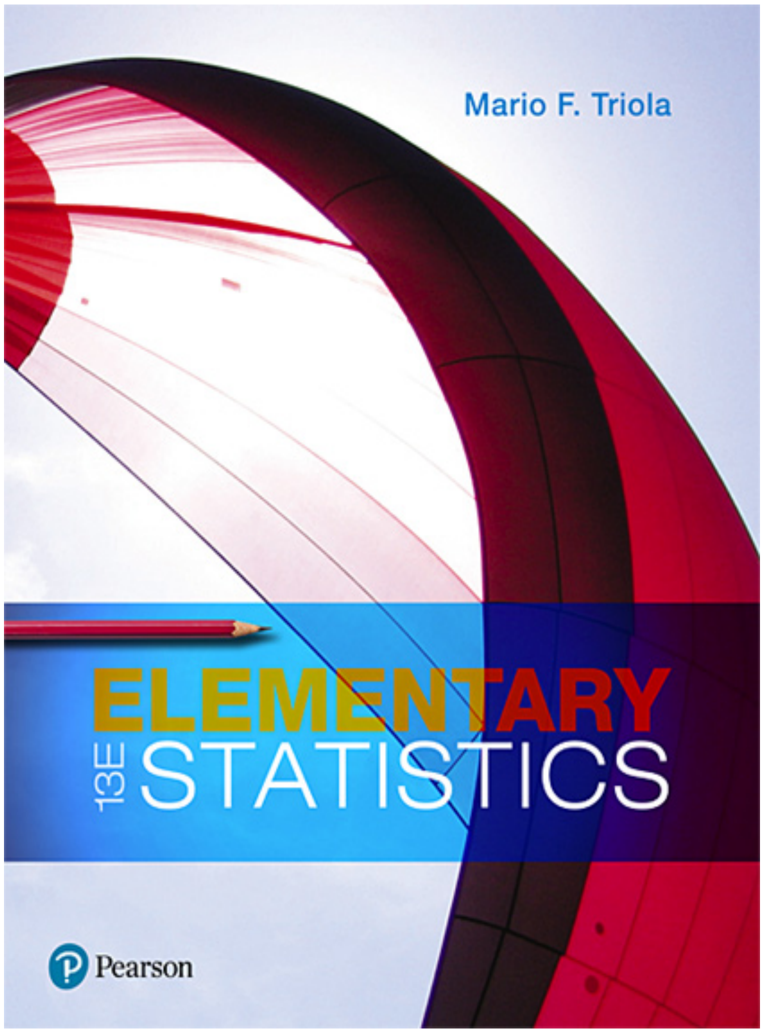
Elementary Statistics

Thirteenth Edition



Chapter 11

Goodness-of-Fit and Contingency Tables

Goodness-of-Fit and Contingency Tables

11-1 Goodness-of-Fit

11-2 Contingency Tables

Key Concept

By "goodness-of-fit" we mean that sample data consisting of observed frequency counts arranged in a single row or column (**called a one-way frequency table**) agree with some particular distribution (such as normal or uniform) being considered. We will use a hypothesis test for the claim that the observed frequency counts agree with the claimed distribution.

Goodness-of-Fit Test

- Goodness-of-Fit Test
  - A **goodness-of-fit test** is used to test the hypothesis that an observed frequency distribution fits (or conforms to) some claimed distribution.

Testing for Goodness-of-Fit: Objective

Conduct a goodness-of-fit test, which is a hypothesis test to determine whether a single row (or column) of frequency counts agrees with some specific distribution (such as uniform or normal).

Testing for Goodness-of-Fit: Notation

O represents the **observed frequency** of an outcome, found from the sample data.

E represents the **expected frequency** of an outcome, found by assuming that the distribution is as claimed.

k represents the **number of different categories** or cells.

n represents the total **number of trials** (or the total of observed sample values).

p represents the **probability** that a sample value falls within a particular category.

Testing for Goodness-of-Fit: Requirements

1. The data have been randomly selected.
2. The sample data consist of frequency counts for each of the different categories.
3. For each category, the **expected** frequency is at least 5. (The expected frequency for a category is the frequency that would occur if the data actually have the distribution that is being claimed. There is no requirement that the **observed** frequency for each category must be at least 5.)

Testing for Goodness-of-Fit: Null and Alternative Hypotheses

H_0: The frequency counts agree with the claimed distribution.

H_1: The frequency counts do not agree with the claimed distribution.

Testing for Goodness-of-Fit: Test Statistic for Goodness-of-Statistic Tests

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

***P*-values:** *P*-values are provided by technology, or a range of *P*-values can be found from Table A-4.

Critical values:

1. Critical values are found in Table A-4 by using $k - 1$ degrees of freedom, where k is the number of categories.
2. Goodness-of-fit hypothesis tests are always **right-tailed**.

Finding Expected Frequencies

Conducting a goodness-of-fit test requires that we identify the **observed** frequencies denoted by *O*, then find the frequencies **expected** (denoted by *E*) with the claimed distribution. There are two different approaches for finding expected frequencies *E*:

- **If the expected frequencies are all equal:**
  Calculate $E = \frac{n}{k}$.
- **If the expected frequencies are not all equal:**
  Calculate $E = np$ **for each individual category.**

Finding Expected Frequencies: Examples (1 of 2)

a. **Equally Likely** A single die is rolled 45 times with the following results. Assuming that the die is fair and all outcomes are equally likely, find the expected frequency *E* for each empty cell.

Outcome	1	2	3	4	5	6
Observed Frequency *O*	13	6	12	9	3	2
Expected Frequency *E*						

With $n = 45$ outcomes and $k = 6$ categories, the expected frequency for each cell is the same: $E = \frac{n}{k} = \frac{45}{6} = 7.5$

If the die is fair and the outcomes are all equally likely, we expect that each outcome should occur about 7.5 times.

Finding Expected Frequencies: Examples (2 of 2)

b. **Not Equally Likely** Using the same results from part (a), suppose that we claim that instead of being fair, the die is loaded so that the outcome of 1 occurs 50% of the time and the other five outcomes occur 10% of the time. The probabilities are listed in the second row below. Using $n = 45$ and the probabilities listed below, we find that for the first cell, $E = np = (45)(0.5) = 22.5$. Each of the other five cells will have the expected value of $E = np = (45)(0.1) = 4.5$.

Outcome	1	2	3	4	5	6
Probability	0.5	0.1	0.1	0.1	0.1	0.1
Observed Frequency O	13	6	12	9	3	2
Expected Frequency E	22.5	4.5	4.5	4.5	4.5	4.5

Measuring Disagreement with the Claimed Distribution

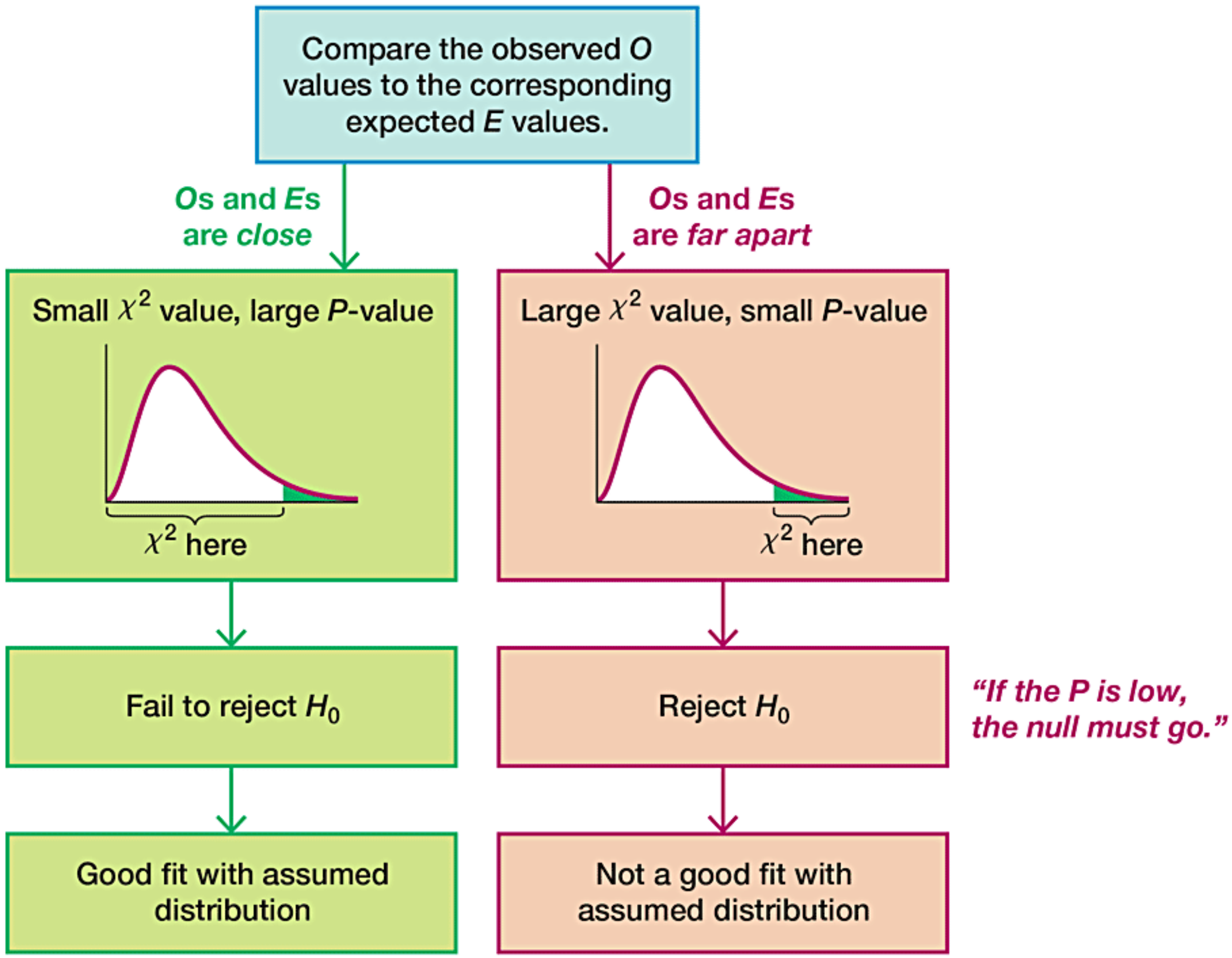
We know that sample frequencies typically differ somewhat from the values we theoretically expect, so we consider the key question:

> **Are the differences between the actual observed frequencies *O* and the theoretically expected frequencies *E* significant?**

To measure the discrepancy between the *O* and *E* values, we use the test statistic

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

Relationships Among the χ^2 Test Statistic, P-value, and Goodness-of-Fit



Example: Last Digits of Weights (1 of 11)

A random sample of 100 weights of Californians is obtained, and the last digits of those weights are summarized in the table (based on data from the California Department of Public Health). When obtaining weights of subjects, it is extremely important to actually measure their weights instead of asking them to report their weights. Test the claim that the sample is from a population of weights in which the last digits do **not** occur with the same frequency.

Last Digits of Weights

Last Digits	Frequency
0	46
1	1
2	2
3	3
4	3
5	30
6	4
7	0
8	8
9	3

Example: Last Digits of Weights (2 of 11)

Solution

REQUIREMENT CHECK (1) The data come from randomly selected subjects. (2) The data do consist of frequency counts, as shown in the table. (3) With 100 sample values and 10 categories that are claimed to be equally likely, each expected frequency is 10, so each expected frequency does satisfy the requirement of being a value of at least 5.

All of the requirements are satisfied.

Example: Last Digits of Weights (3 of 11)

Solution

Step 1: The original claim is that the digits do not occur with the same frequency. That is, at least one of the probabilities, p_0, p_1, . . . , p_9, is different from the others.

Step 2: If the original claim is false, then all of the probabilities are the same.

That is, $p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$.

Example: Last Digits of Weights (4 of 11)

Solution

Step 3: The null hypothesis must contain the condition of equality, so we have

H_0: $p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$

H_1: At least one of the probabilities is different from the others.

Step 4: No significance level was specified, so we select the common choice of $\alpha = 0.05$.

Example: Last Digits of Weights (5 of 11)

Solution

Step 5: Because we are testing a claim about the distribution of the last digits being a uniform distribution, we use the goodness-of-fit test described in this section. The χ^2 distribution is used.

Example: Last Digits of Weights (6 of 11)

Solution

Step 6: The observed frequencies *O* are in the table. Each corresponding expected frequency *E* is equal to 10. The Excel add-in XLSTAT is used to obtain the results. The table on the next slide shows the manual computation of the χ^2 test statistic.

Last Digits of Weights

Last Digits	Frequency
0	46
1	1
2	2
3	3
4	3
5	30
6	4
7	0
8	8
9	3

XLSTAT

Chi-square (Observed value)	212.8000
Chi-square (Critical value)	16.9190
DF	9
p-value	< 0.0001
alpha	0.05

Example: Last Digits of Weights (7 of 11)

Solution

Last Digit	Observed Frequency O	Expected Frequency E	$O - E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	46	10	36	1296	129.6
1	1	10	−9	81	8.1
2	2	10	−8	64	6.4
3	3	10	−7	49	4.9
4	3	10	−7	49	4.9
5	30	10	20	400	40.0
6	4	10	−6	36	3.6
7	0	10	−10	100	10.0
8	8	10	−2	4	0.4
9	3	10	−7	49	4.9

$$X^2 = \sum \frac{(O-E)^2}{E} = 212.8$$

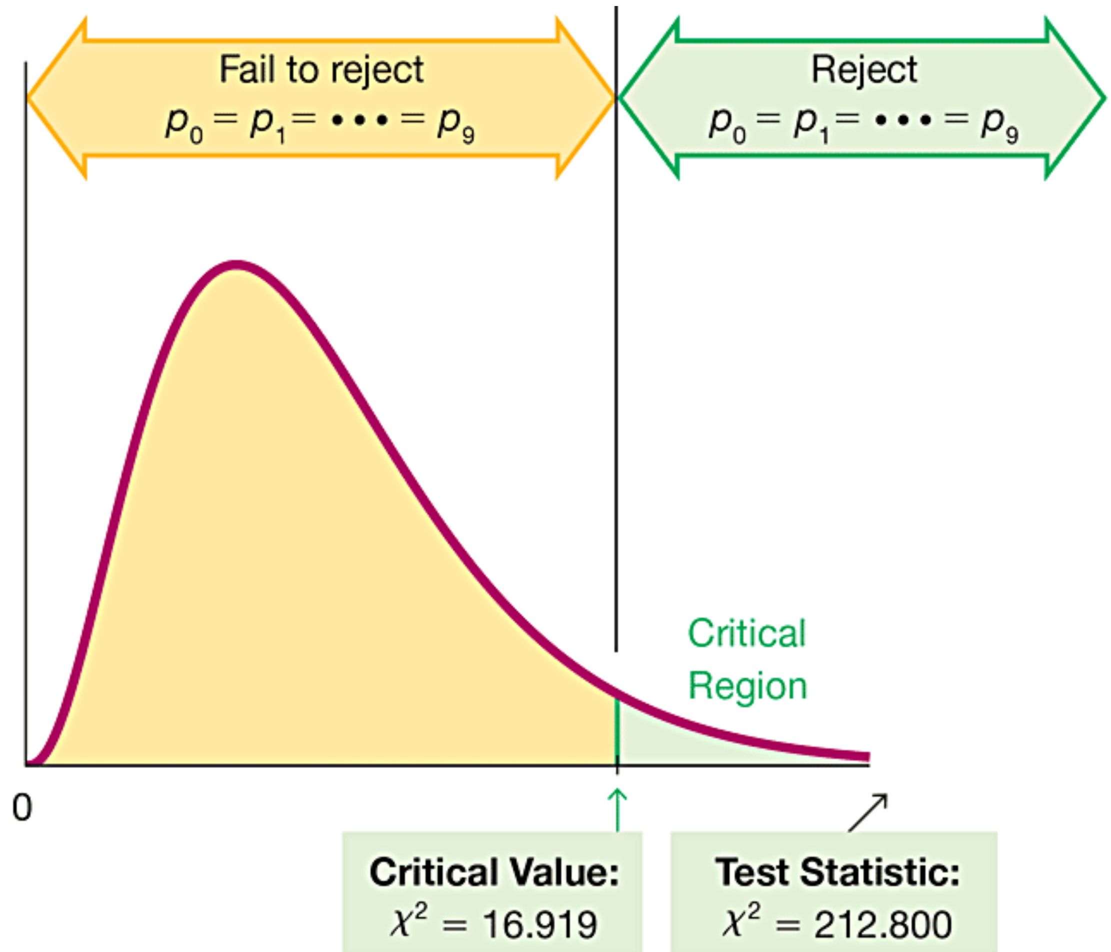
Example: Last Digits of Weights (8 of 11)

Solution

Step 6 (con't): The test statistic is $\chi^2 = 212.800$. The critical value is $\chi^2 = 16.919$ (found in Table A-4 with $\alpha = 0.05$ in the right tail and degrees of freedom equal to $k - 1 = 9$). The P-value is less than 0.0001. The test statistic and critical value are shown on the next slide.

Example: Last Digits of Weights (9 of 11)

Solution



Example: Last Digits of Weights (10 of 11)

Solution

Step 7: If we use the *P*-value method of testing hypotheses, we see that the *P*-value is small (less than 0.0001), so we reject the null hypothesis. If we use the critical value method of testing hypotheses, the previous figure shows that the test statistic falls in the critical region, so there is sufficient evidence to reject the null hypothesis.

Step 8: There is sufficient evidence to support the claim that the last digits do not occur with the same relative frequency.

Example: Last Digits of Weights

Interpretation

This goodness-of-fit test suggests that the last digits do not provide a good fit with the claimed uniform distribution of equally likely frequencies. Instead of actually weighing the subjects, it appears that the subjects reported their weights. In fact, the weights are from the California Health Interview Survey (CHIS), and the title of that survey indicates that subjects were interviewed, not measured. Because those weights are reported, the reliability of the data is very questionable.

Example: Benford's Law – Detecting Computer Intrusions (1 of 9)

According to **Benford's law**, many data sets have the property that the leading (first) digits follow the distribution shown in the first two rows of the table below. The bottom row lists the frequencies of leading digits of Internet traffic inter-arrival times. Do the frequencies in the bottom row fit the distribution according to Benford's Law?

Leading Digit	1	2	3	4	5	6	7	8	9
Benford's Law: Distribution of Leading Digits	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%
Sample 2 of Leading Digits	69	40	42	26	25	16	16	17	20

Example: Benford's Law – Detecting Computer Intrusions

Solution

REQUIREMENT CHECK (1) The sample data are randomly selected from a larger population. (2) The sample data do consist of frequency counts. (3) Each expected frequency is at least 5. The lowest expected frequency is $271 \cdot 0.46 = 12.466$. All of the requirements are satisfied.

Example: Benford's Law – Detecting Computer Intrusions (3 of 9)

Solution

Step 1: The original claim is that the leading digits fit the distribution given as Benford's law. Using subscripts corresponding to the leading digits, we can express this claim as $p_1 = 0.301$ and $p_2 = 0.176$ and $p_3 = 0.125$ and … and $p_9 = 0.046$.

Step 2: If the original claim is false, then at least one of the proportions does not have the value as claimed.

Example: Benford's Law – Detecting Computer Intrusions (4 of 9)

Solution

Step 3: The null hypothesis must contain the condition of equality, so we have

H_0: $p_1 = 0.301$ and $p_2 = 0.176$ and $p_3 = 0.125$ and … and $p_9 = 0.046$.

H_1: At least one of the proportions is not equal to the given claimed value.

Step 4: The significance level is not specified, so we use the common choice of $\alpha = 0.05$.

Example: Benford's Law – Detecting Computer Intrusions (5 of 9)

Solution

Step 5: Because we are testing a claim that the distribution of leading digits fits the distribution given by Benford's law, we use the goodness-of-fit test described in this section.

Step 6: The table shows the calculations of the components of the χ^2 test statistic for the leading digits of 1 and 2.

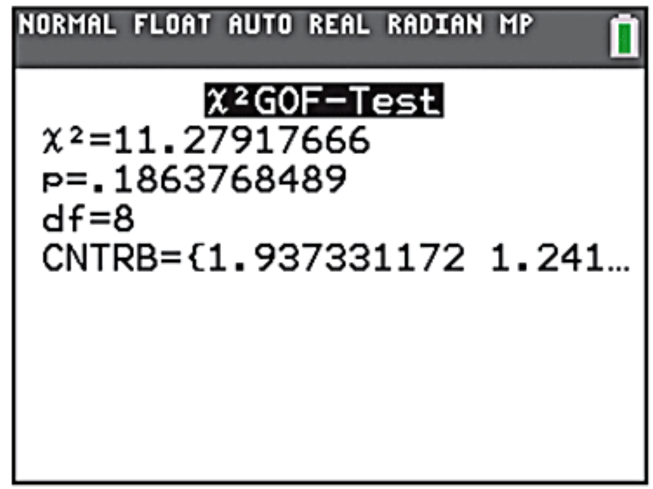
Leading Digit	Observed Frequency O	Expected Frequency $E = np$	$O - E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1	69	$271 \cdot 0.301 = 81.5710$	−12.5710	158.0300	1.9373
2	40	$271 \cdot 0.176 = 47.6960$	−7.6960	59.2284	1.2418

Example: Benford's Law – Detecting Computer Intrusions (6 of 9)

Solution

Step 6 (con't): If we include all nine leading digits, we get the test statistic of $\chi^2 = 11.2792$, as shown in the TI-84 Plus C calculator display. The critical value is $\chi^2 = 15.507$ (found in Table A-4, with $\alpha = 0.05$ in the right tail and degrees of freedom equal to $k - 1 = 8$). The TI-84 Plus C calculator display shows the value of the test statistic as well as the *P*-value of 0.186.

TI-84 Plus C



Pearson

Example: Benford's Law – Detecting Computer Intrusions (7 of 9)

Solution

Step 7: The *P*-value of 0.186 is greater than the significance level of 0.05, so there is not sufficient evidence to reject the null hypothesis.

Step 8: There is not sufficient evidence to warrant rejection of the claim that the 271 leading digits fit the distribution given by Benford's law.

Example: Benford’s Law – Detecting Computer Intrusions (8 of 9)

Interpretation

The sample of leading digits does not provide enough evidence to conclude that the Benford’s law distribution is not being followed. There is not sufficient evidence to support a conclusion that the leading digits are from inter-arrival times that are not from normal traffic, so there is not sufficient evidence to conclude that an Internet intrusion has occurred.

Example: Benford's Law – Detecting Computer Intrusions (9 of 9)

In the figure we use a green line to graph the expected proportions given by Benford's law along with a red line for the observed proportions. The figure allows us to visualize the "goodness-of-fit" between the distribution given by Benford's law and the frequencies that were observed. The green and red lines agree reasonably well, so it appears that the observed data fit the expected values reasonably well.