The Chi-Square Test
Sections 25.4, 25.5, 25.6
Lecture 46

Robb T. Koether

Hampden-Sydney College

Wed, Apr 13, 2016

Outline

Outline

Two-Way Tables

- We have two categorical variables, e.g., age group and cell-phone type.
- We gather count data for every combination of values for the two variables.

	Landline	Cell	Both	Neither	Total
18 - 30	50	180	20	50	300
31 - 50	100	150	100	50	400
50 - 69	150	150	150	50	500
Total	300	500	250	150	1200

Expected Counts

- Compute the expected counts using the formula

$$\text{Expected count} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}.$$

	Landline	Cell	Both	Neither	Total
18 - 30	50	180	20	50	300
31 - 50	100	150	100	50	400
50 - 69	150	150	150	50	500
Total	300	480	270	150	1200

Expected Counts

- Compute the expected counts using the formula

$$\text{Expected count} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}.$$

	Landline	Cell	Both	Neither	Total
18 - 30	50 75.0	180 120.0	20 67.5	50 37.5	300
31 - 50	100 100.0	150 160.0	100 90.0	50 50.0	400
50 - 69	150 125.0	150 200.0	150 112.5	50 62.5	500
Total	300	480	270	150	1200

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\chi^2 = \frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5}$$
$$\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50}$$
$$\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\ 8.333\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\ 8.333 + 30.0\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\ 8.333 + 30.0 + 33.426\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\ 8.333 + 30.0 + 33.426 + 4.167\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\, 8.333 + 30.0 + 33.426 + 4.167 \\ &+ 0.0\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\ 8.333 + 30.0 + 33.426 + 4.167 \\ &+ 0.0 + 0.625\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\, 8.333 + 30.0 + 33.426 + 4.167 \\ &+ 0.0 + 0.625 + 1.111\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\, 8.333 + 30.0 + 33.426 + 4.167 \\ &+ 0.0 + 0.625 + 1.111 + 0.0\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}\chi^2 = &\frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\ &\frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\ &\frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\ = &\, 8.333 + 30.0 + 33.426 + 4.167 \\ &+ 0.0 + 0.625 + 1.111 + 0.0 \\ &+ 5.0\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}
\chi^2 = & \frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\
& \frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\
& \frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\
= & 8.333 + 30.0 + 33.426 + 4.167 \\
& + 0.0 + 0.625 + 1.111 + 0.0 \\
& + 5.0 + 12.5
\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}
\chi^2 = & \frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\
& \frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\
& \frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\
= & 8.333 + 30.0 + 33.426 + 4.167 \\
& + 0.0 + 0.625 + 1.111 + 0.0 \\
& + 5.0 + 12.5 + 12.5
\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}
\chi^2 = & \frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\
& \frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\
& \frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\
= & \, 8.333 + 30.0 + 33.426 + 4.167 \\
& + 0.0 + 0.625 + 1.111 + 0.0 \\
& + 5.0 + 12.5 + 12.5 + 2.5
\end{aligned}$$

The χ^2 Statistic

- Compute the χ^2 statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

$$\begin{aligned}
\chi^2 = & \frac{(50-75.0)^2}{75.0} + \frac{(180-120.0)^2}{120.0} + \frac{(20-67.5)^2}{67.5} + \frac{(50-37.5)^2}{37.5} \\
& \frac{(100-100)^2}{100} + \frac{(150-160)^2}{160} + \frac{(100-90)^2}{90} + \frac{(50-50)^2}{50} \\
& \frac{(250-125)^2}{125} + \frac{(150-200)^2}{200} + \frac{(150-112.5)^2}{112.5} + \frac{(50-62.5)^2}{62.5} \\
= & \, 8.333 + 30.0 + 33.426 + 4.167 \\
& + 0.0 + 0.625 + 1.111 + 0.0 \\
& + 5.0 + 12.5 + 12.5 + 2.5 \\
= & \, 110.162.
\end{aligned}$$

Outline

The chi^2 Test

- Our procedure will follow the same 6 steps as always.
  1. State the hypotheses.
  2. Give the value of α.
  3. Write the formula for the test statistic.
  4. Calculate the value of the test statistic and show the expected counts.
  5. Calculate the *p*-value.
  6. Draw a conclusion.

The chi^2 Test

- The null hypothesis says that there is no difference in the distributions among the rows or among the columns.
- That is, the two variables are independent.

$$H_0: \text{The variables are independent}$$

- The alternative hypothesis says the opposite.

$$H_a: \text{The variables are not independent}$$

The *chi*^{2} Test

- The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

- The degrees of freedom is

$$df = (\text{No. of rows} - 1) \times (\text{No. of columns} - 1).$$

- To find the *p*-value of χ^2, use the $\chi^2\texttt{cdf}$ function on the TI-83.

Outline

The χ^2 Test

Example (Computing χ^2)

- Test whether a person's age group and cell-phone type are independent.

	Landline	Cell	Both	Neither	Total
18 - 30	50	180	20	50	300
31 - 50	100	150	100	50	400
50 - 69	150	150	150	50	500
Total	300	480	270	150	1200

The χ^2 Test

Example (Computing χ^2)

- Test whether a person's age group and cell-phone type are independent.

	Landline	Cell	Both	Neither	Total
18 - 30	50 75.0	180 120.0	20 67.5	50 37.5	300
31 - 50	100 100.0	150 160.0	100 90.0	50 50.0	400
50 - 69	150 125.0	150 200.0	150 112.5	50 62.5	500
Total	300	480	270	150	1200

The χ^2 Test

Example (Computing χ^2)

(1)

H_0: The variables are independent

H_a: The variables are not independent

(2) Let $\alpha = 0.05$.

The χ^2 Test

Example (Computing χ^2)

(3) The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

(4) We calculate $\chi^2 = 110.162$. The number of degrees of freedom is $2 \times 3 = 6$.

The χ^2 Test

Example (Computing χ^2)

(3) The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

(4) We calculate $\chi^2 = 110.162$. The number of degrees of freedom is $2 \times 3 = 6$. (Thus, χ^2 has mean 6 and standard deviation $\sqrt{12} = 3.46$.)

The χ^2 Test

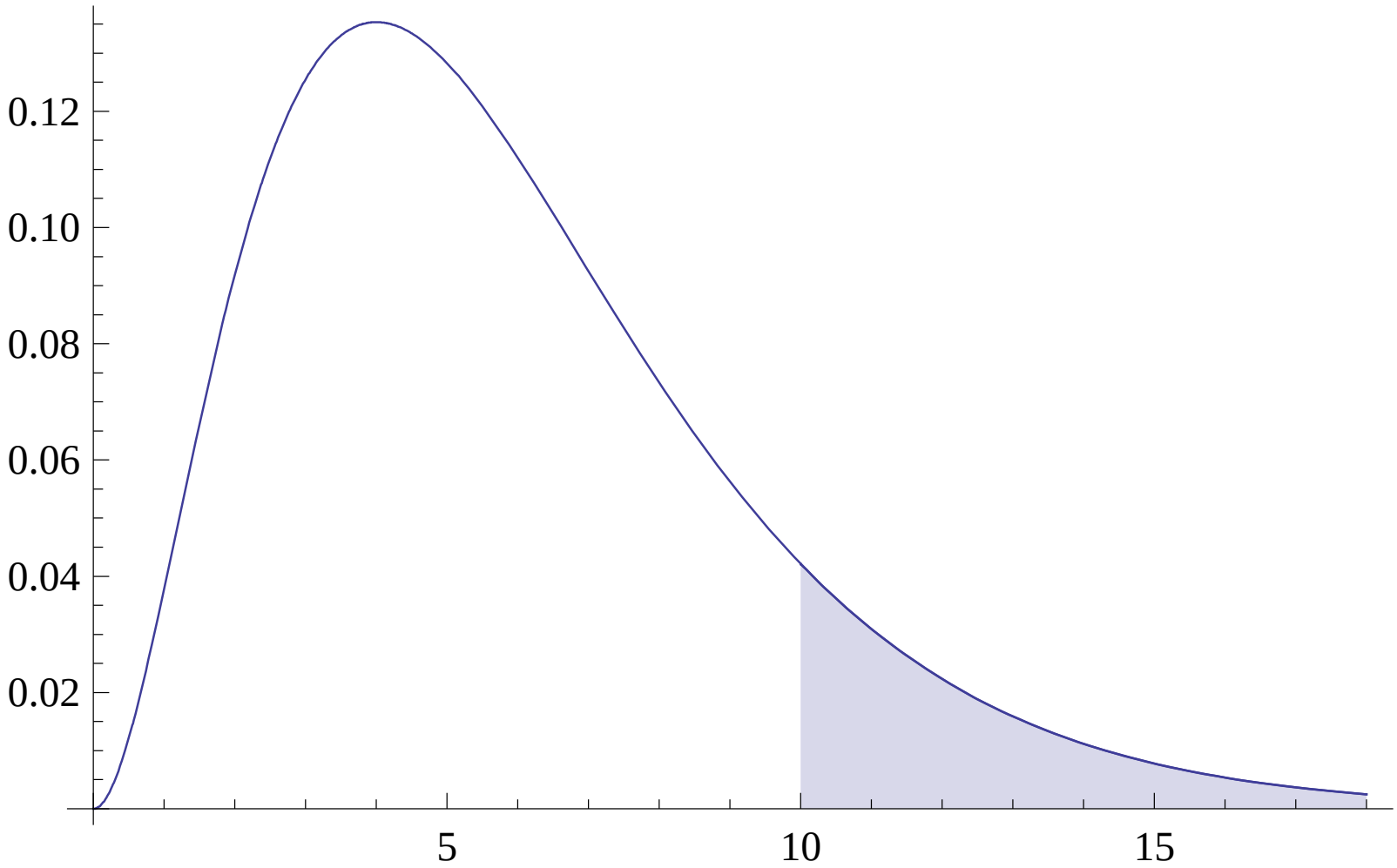
Example (Computing χ^2)

(5) The *p*-value is

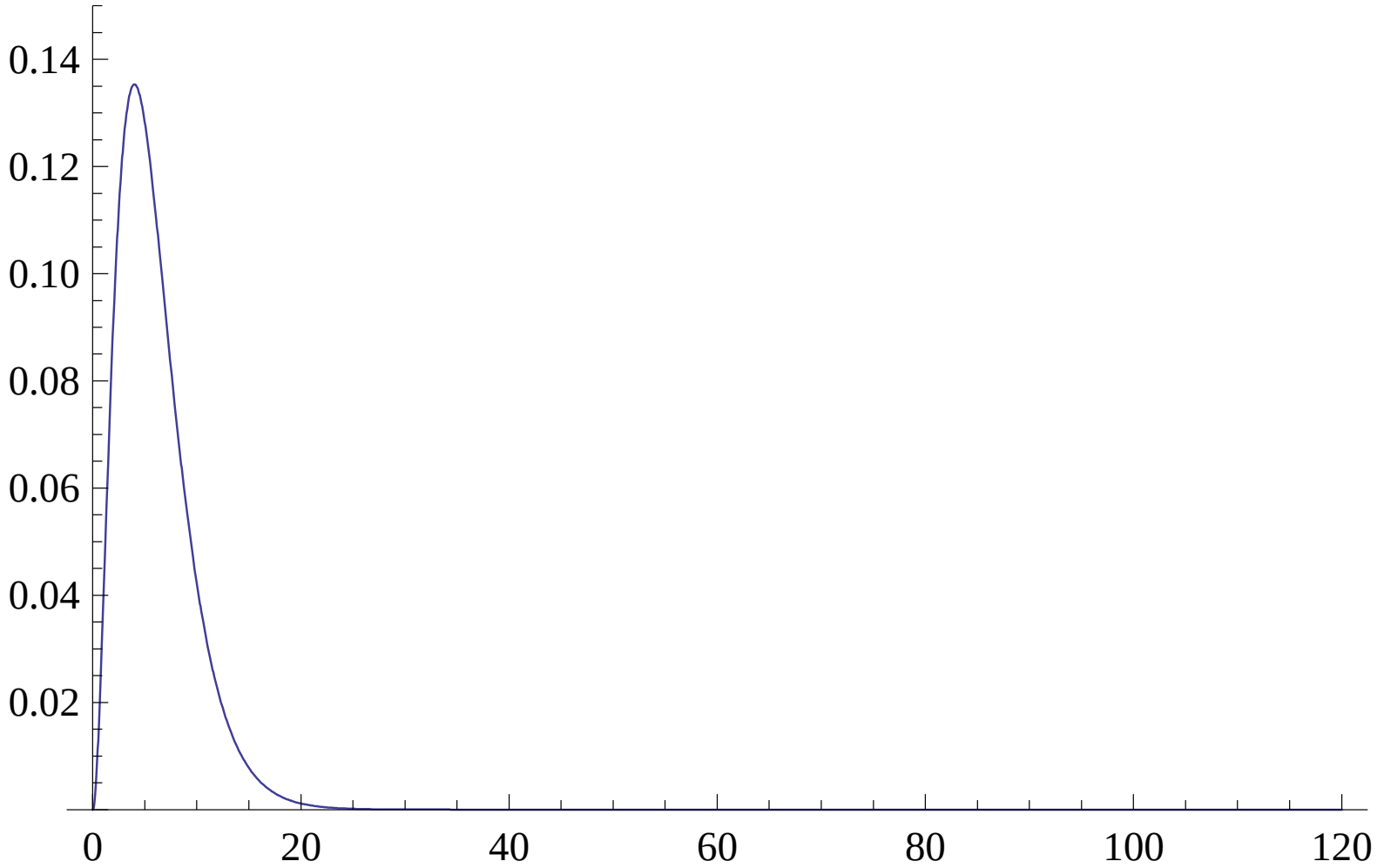
$$\begin{aligned} p\text{-value} &= \chi^2\texttt{cdf(110.162,E99,6)} \\ &= 1.885 \times 10^{-21}. \end{aligned}$$

(6) Reject H_0 and conclude that age group and cell-phone type are not independent.

The χ^2 Test



The χ^2 Test



The χ^2 Test

Example (Computing χ^2)

- Test whether a person's political party affiliation and state of residence are independent.

	KS	NE	SD	ND	Total
Republican	470	290	120	110	1000
Democrat	255	185	90	80	600
Total	725	475	210	190	1600

The χ^2 Test

Example (Computing χ^2)

- Test whether a person's political party affiliation and state of residence are independent.

	KS	NE	SD	ND	Total
Republican	470 488.59	290 293.91	120 129.94	110 117.56	1000
Democrat	255 276.41	185 181.09	90 80.06	80 72.44	600
Total	725	475	210	190	1600

The χ^2 Test

Example (Computing χ^2)

(1)

H_0: The variables are independent

H_a: The variables are not independent

(2) Let $\alpha = 0.05$.

The χ^2 Test

Example (Computing χ^2)

(3) The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

(4) We calculate $\chi^2 = 6.0849$. The number of degrees of freedom is $1 \times 3 = 3$.

The χ^2 Test

Example (Computing χ^2)

(3) The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$$

(4) We calculate $\chi^2 = 6.0849$. The number of degrees of freedom is $1 \times 3 = 3$. (Thus, χ^2 has mean 3 and standard deviation $\sqrt{6} = 2.449$.)

The χ^2 Test

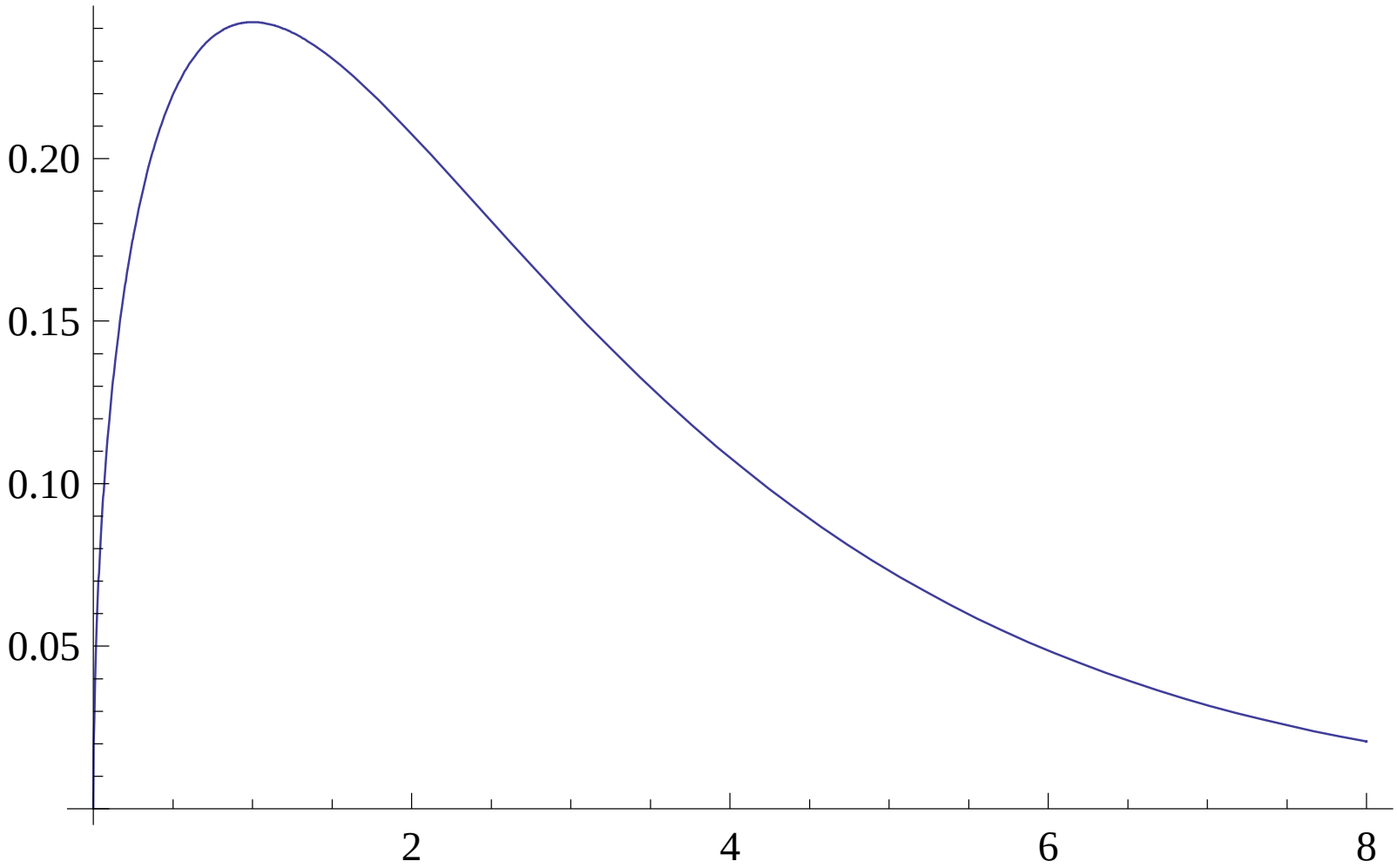
Example (Computing χ^2)

(5) The *p*-value is

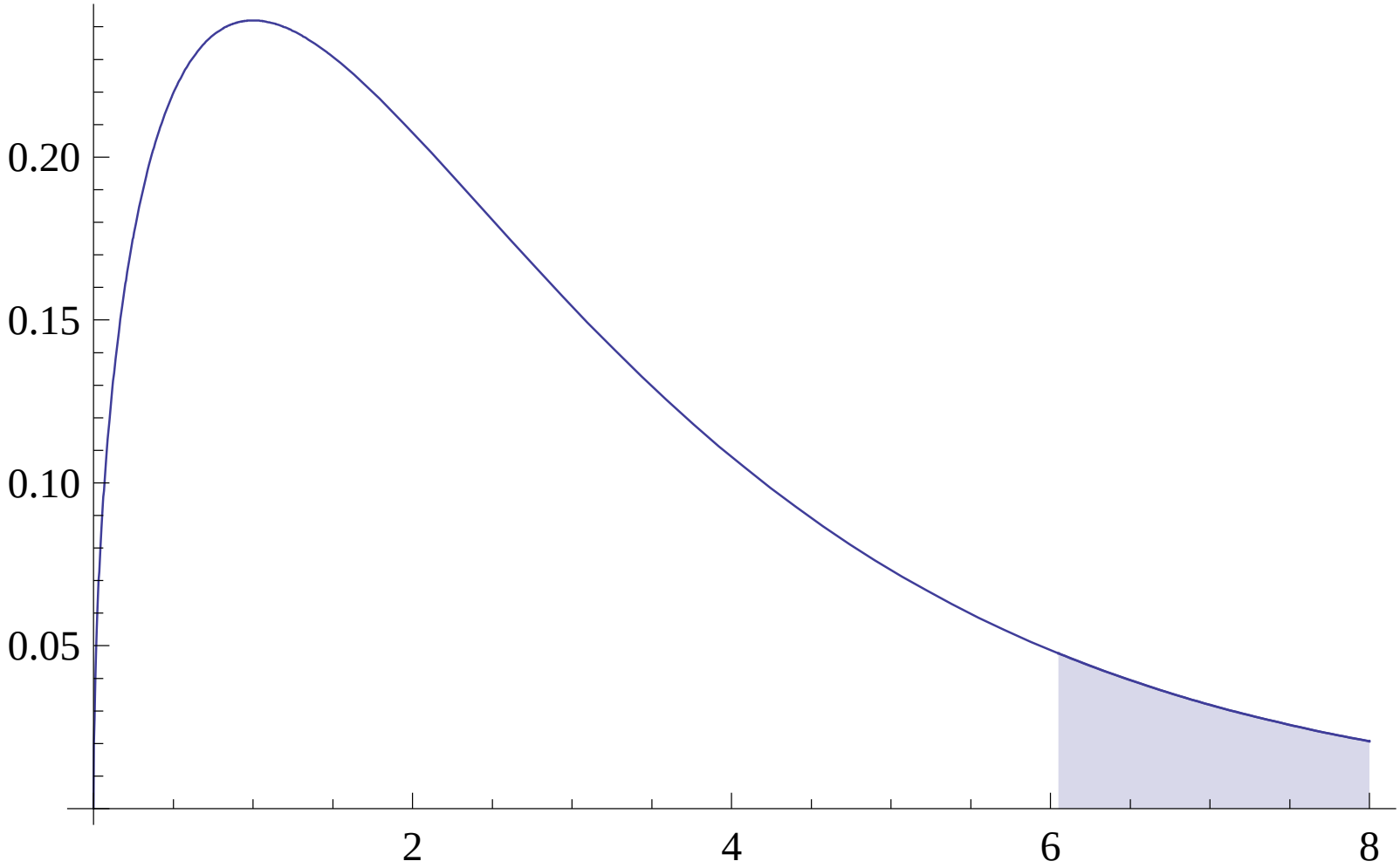
$$\begin{aligned} p\text{-value} &= \chi^2\texttt{cdf(6.0849,E99,3)} \\ &= 0.1075. \end{aligned}$$

(6) We cannot reject H_0. We conclude that party affiliation and state of residence are independent.

The χ^2 Test



The χ^2 Test



Outline

Assignment

Assignment

- Read Sections 25.4, 25.5, 25.6.
- Apply Your Knowledge: 7, 8, 9.
- Check Your Skills: 28.
- Exercises 36, 39, 40, 42.