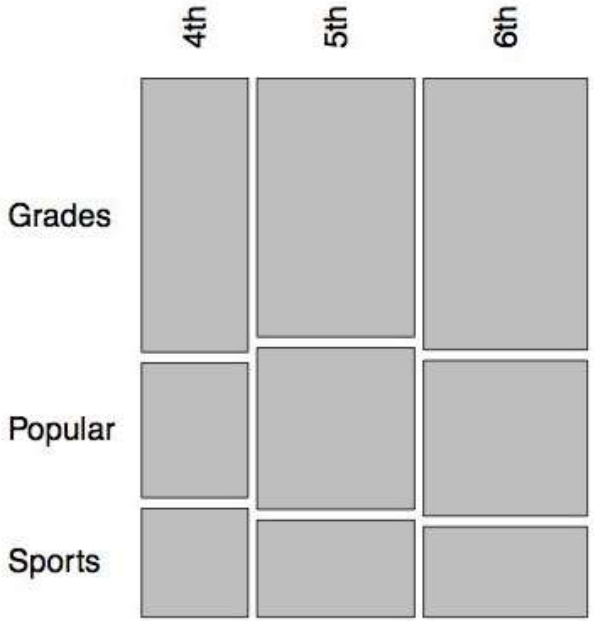
Chi-Square Test of Independence

Popular kids

In the dataset `popular`, students in grades 4-6 were asked whether good grades, athletic ability, or popularity was most important to them. A two-way table separating the students by grade and by choice of most important factor is shown below. Do these data provide evidence to suggest that goals vary by grade?

	Grades	Popular	Sports
4^{th}	63	31	25
5^{th}	88	55	33
6^{th}	96	55	32



Chi-square test of independence

- The hypotheses are:
  H_0: Grade and goals are independent. Goals do not vary by grade.
  H_A: Grade and goals are dependent. Goals vary by grade.

Chi-square test of independence

- The hypotheses are:
  H_0: Grade and goals are independent. Goals do not vary by grade.
  H_A: Grade and goals are dependent. Goals vary by grade.

- The test statistic is calculated as

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O-E)^2}{E} \quad \text{where} \quad df = (R-1) \times (C-1),$$

  where k is the number of cells, R is the number of rows, and C is the number of columns.

Note: *we calculate df differently for one-way and two-way tables.*

Chi-square test of independence

- The hypotheses are:
  H_0: Grade and goals are independent. Goals do not vary by grade.
  H_A: Grade and goals are dependent. Goals vary by grade.

- The test statistic is calculated as

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O-E)^2}{E} \quad \text{where} \quad df = (R-1)\times(C-1),$$

  where k is the number of cells, R is the number of rows, and C is the number of columns.

Note: *we calculate df differently for one-way and two-way tables.*

- The p-value is the area under the χ^2_{df} curve, above the calculated test statistic.

Expected counts in two-way tables

$$\text{Expected Count} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

Expected counts in two-way tables

$$\text{Expected Count} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

Expected counts in two-way tables

$$\text{Expected Count} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

$$E_{row\ 1, col\ 1} = \frac{119 \times 247}{478} = 61$$

Expected counts in two-way tables

$$\text{Expected Count} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

$$E_{row\ 1,col\ 1} = \frac{119 \times 247}{478} = 61 \qquad E_{row\ 1,col\ 2} = \frac{119 \times 141}{478} = 35$$

Expected counts in two-way tables

What is the expected count for the highlighted cell?

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	*55*	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

(a) 176 x 141 / 478
(b) 119 x 141 / 478
(c) 176 x 247 / 478
(d) 176 x 478 / 478

Expected counts in two-way tables

What is the expected count for the highlighted cell?

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	*55*	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

(a) 176 x 141 / 478 → 52

more than expected # of 5th graders have a goal of being popular

(b) 119 x 141 / 478

(c) 176 x 247 / 478

(d) 176 x 478 / 478

Calculating the test statistic in two-way tables

Expected counts are shown in blue next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 *61*	31 *35*	25 *23*	119
5^{th}	88 *91*	55 *52*	33 *33*	176
6^{th}	96 *95*	55 *54*	32 *34*	183
Total	247	141	90	478

Calculating the test statistic in two-way tables

Expected counts are shown in blue next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 *61*	31 *35*	25 *23*	119
5^{th}	88 *91*	55 *52*	33 *33*	176
6^{th}	96 *95*	55 *54*	32 *34*	183
Total	247	141	90	478

$$\chi^2 = \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121$$

Calculating the test statistic in two-way tables

Expected counts are shown in blue next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 *61*	31 *35*	25 *23*	119
5^{th}	88 *91*	55 *52*	33 *33*	176
6^{th}	96 *95*	55 *54*	32 *34*	183
Total	247	141	90	478

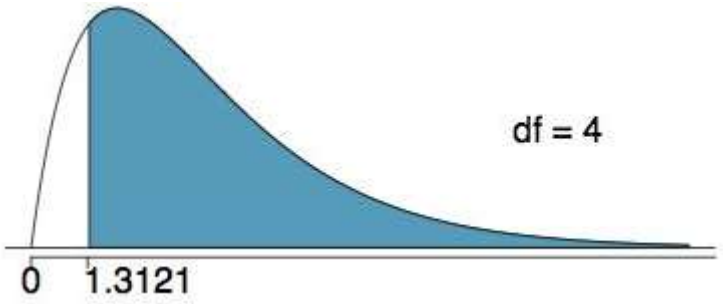
$$\chi^2 = \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121$$

$$df = (R-1) \times (C-1) = (3-1) \times (3-1) = 2 \times 2 = 4$$

Calculating the p-value

Which of the following is the correct p-value for this hypothesis test?

$\chi^2_{df} = 1.3121$ df = 4

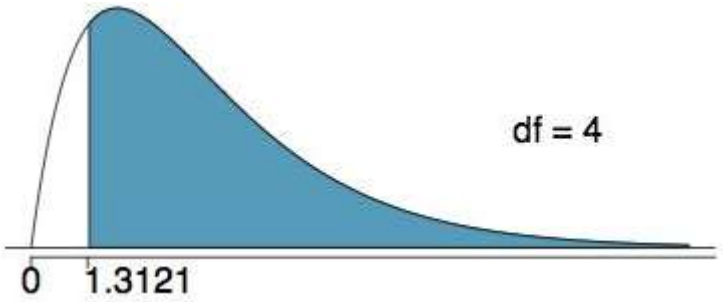


(a) more than 0.3
(b) between 0.3 and 0.2
(c) between 0.2 and 0.1
(d) between 0.1 and 0.05
(e) less than 0.001

Calculating the p-value

Which of the following is the correct p-value for this hypothesis test?

$\chi^2_{df} = 1.3121$ $df = 4$



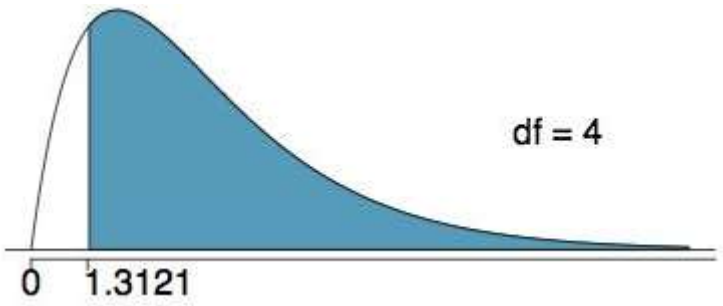
(a) more than 0.3
(b) between 0.3 and 0.2
(c) between 0.2 and 0.1
(d) between 0.1 and 0.05
(e) less than 0.001

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52

Calculating the p-value

Which of the following is the correct p-value for this hypothesis test?

$\chi^2_{df} = 1.3121$ $\quad\quad$ df = 4



(a) more than 0.3

(b) between 0.3 and 0.2

(c) between 0.2 and 0.1

(d) between 0.1 and 0.05

(e) less than 0.001

Conclusion

Do these data provide evidence to suggest that goals vary by grade?

H_0: Grade and goals are independent.
Goals do not vary by grade.

H_A: Grade and goals are dependent.
Goals vary by grade.

Conclusion

Do these data provide evidence to suggest that goals vary by grade?

H_0: Grade and goals are independent.
Goals do not vary by grade.

H_A: Grade and goals are dependent.
Goals vary by grade.

Since the p-value is large, we fail to reject H_0.The data do not provide convincing evidence that grade and goals are dependent. It doesn't appear that goals vary by grade.