

Section 6.4

Chi-Square Test of Independence

Stats 7 Summer Session II 2022

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.

	Grades	Popular	Sports
<i>4th</i>	63	31	25
<i>5th</i>	88	55	33
<i>6th</i>	96	55	32

	4th	5th	6th
Grades			
Popular			
Sports			

Do these data provide evidence to suggest that 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

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

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

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

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

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

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

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) $176 \times 141 / 478$

(b) $119 \times 141 / 478$

(c) $176 \times 247 / 478$

(d) $176 \times 478 / 478$

→ 52

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

Calculating the test statistic in two-way tables

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

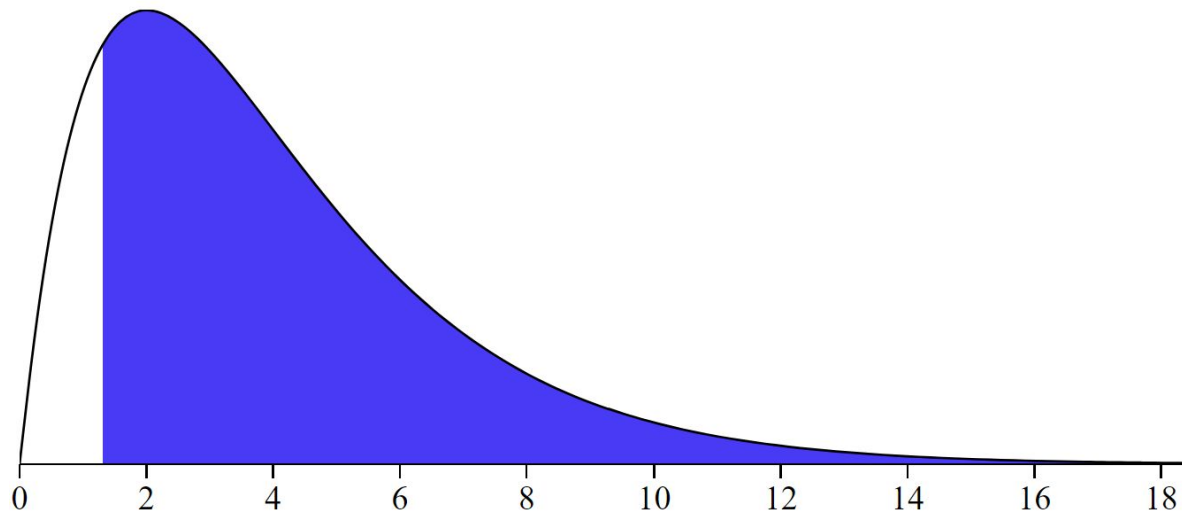
	Grades	Popular	Sports	Total
4 th	63 61	31 35	25 23	119
5 th	88 91	55 52	33 33	176
6 th	96 95	55 54	32 34	183
Total	247	141	90	478

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

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

Calculating the p-value

Calculate the p-value for the area above 1.3121, under the chi square distribution with 4 degrees of freedom.



```
> 1 - pchisq(1.3121, df = 4)  
[1] 0.8593193
```

We have a 85.93% probability of observing data as or more extreme from what we observed, if grade and goals are truly independent.

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 (p-value = 0.8593) that grade and goals are dependent. It doesn't appear that goals vary by grade.

Derivative of slides developed by Mine Çetinkaya-Rundel of OpenIntro.
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.
The slides may be copied, edited, and/or shared via the
[CC BY-SA license](#)