10.2 Hypothesis Testing with Two-Way Tables

Part 2: more examples

3x3 Two way table

2x3 Two-way table (worksheet)

Example 2:

- Is there an association between the type of school area and the students' choice of good grades, athletic ability, or popularity as most important?

- Two categorical variables:

  School area (Rural, Suburban, Urban)

  Goals (Grades, Popularity, Sports)

- 335- 4th, 5th, and 6th graders were asked, and their responses were recorded. The information is shown in the two-way table below:

			Goals	
		Grades	Popularity	Sports
	Rural	57	50	42
School area	Suburban	87	42	22
	Urban	24	6	5

Hypothesis test for two-way tables

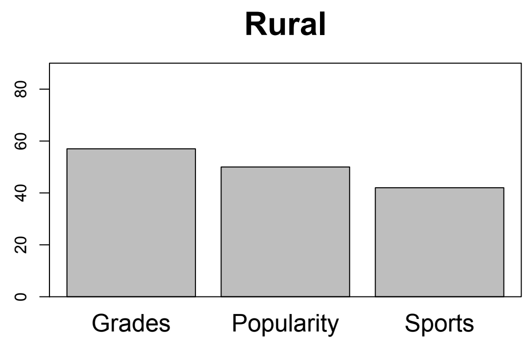
- If there IS NOT a relationship, then the categorical variables do not impact each other.
  - H_0: the variables are independent (no relationship exists)
- If there IS a relationship, then the categorical variables DO impact each other.
  - H_a: there is a relationship between the two variables

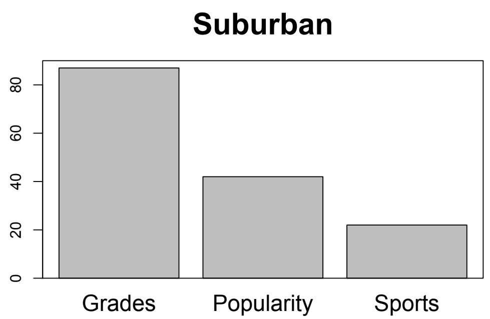
Hypotheses in the context of the data

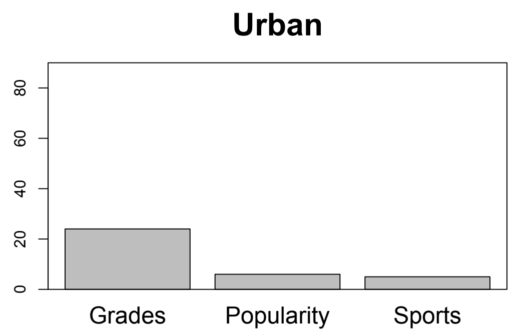
- H_0: The location of school **has no bearing** on what activities students consider to be most worth their time.

- H_a: The location of school **does impact** what activities students consider to be most worth their time.

Observed counts

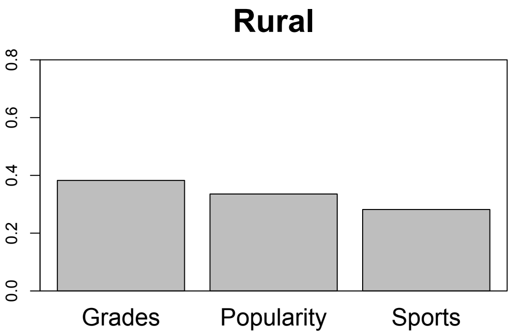


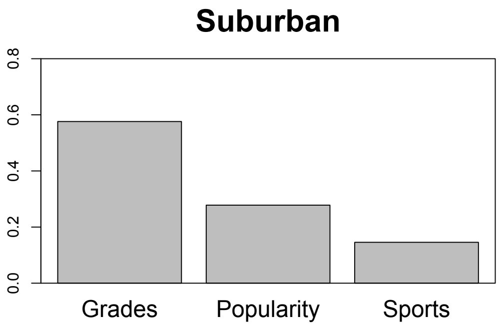


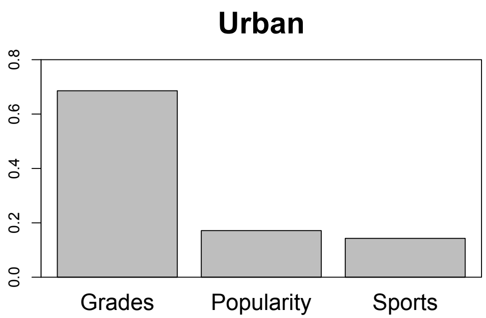


Fewer Urban kids were surveyed

As relative frequencies within each school type





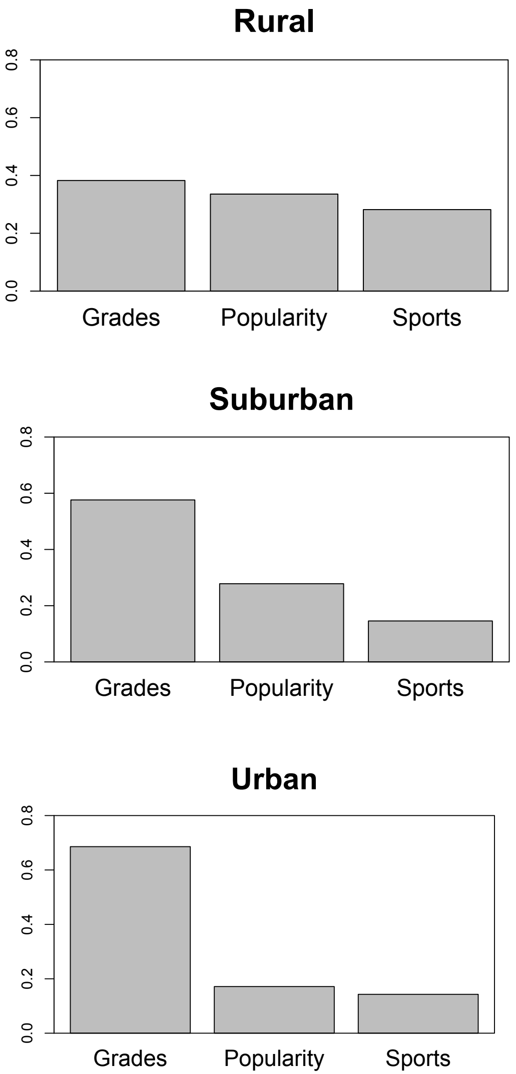


As relative frequencies within each school type

Are these three probability distributions similar enough to suggest that school type does not affect what kids think is most important?

If so, accept H_0.

If not, reject H_0.



Performing the hypothesis test

- We need row totals and column totals.

	Grades	Popularity	Sports	totals
Rural	57	50	42	149
Suburban	87	42	22	151
Urban	24	6	5	35
totals	168	98	69	335

Performing the hypothesis test

- Let's calculate the frequencies (counts) we would have expected if the null were true (i.e. there the variables were independent).

	Grades	Popularity	Sports	totals
Rural	?	?	?	149
Suburban	?	?	?	151
Urban	?	?	?	35
totals	168	98	69	335

Performing the hypothesis test

- We convert the row and column totals to **relative frequencies**…

	Grades	Popularity	Sports	totals
Rural	?	?	?	149/335
Suburban	?	?	?	151/335
Urban	?	?	?	35/335
totals	168/335	98/335	69/335	335/335

Performing the hypothesis test

- We convert the row and column totals to **relative frequencies**…

	Grades	Popularity	Sports	totals
Rural	?	?	?	0.445
Suburban	?	?	?	0.451
Urban	?	?	?	0.104
totals	0.501	0.293	0.206	1

Performing the hypothesis test

- If the variables are independent (i.e. H_0 is true), then…

$$\begin{aligned} P(\text{Rural and Grades}) &= P(\text{Rural}) \times P(\text{Grades}) \\ &= 0.445 \times 0.501 \\ &= 0.223 \end{aligned}$$

Relative frequency table

	Grades	Popularity	Sports	totals
Rural	0.223	?	?	0.445
Suburban	?	?	?	0.451
Urban	?	?	?	0.104
totals	0.501	0.293	0.206	1

Performing the hypothesis test

- If the variables are independent (i.e. H_0 is true), then…

$$
\begin{aligned}
P(\text{Rural and Poplr}) &= P(\text{Rural}) \times P(\text{Poplr}) \\
&= 0.445 \times 0.293 \\
&= 0.130
\end{aligned}
$$

Relative frequency table

	Grades	Popularity	Sports	totals
Rural	0.223	0.130	?	0.445
Suburban	?	?	?	0.451
Urban	?	?	?	0.104
totals	0.501	0.293	0.206	1

Performing the hypothesis test

- If the variables are independent (i.e. H_0 is true), then…

$$\begin{aligned} P(\text{Rural and Sports}) &= P(\text{Rural}) \times P(\text{Sports}) \\ &= 0.445 \times 0.206 \\ &= 0.092 \end{aligned}$$

Relative frequency table

	Grades	Popularity	Sports	totals
Rural	0.223	0.130	0.092	0.445
Suburban	?	?	?	0.451
Urban	?	?	?	0.104
totals	0.501	0.293	0.206	1

Performing the hypothesis test

- Filling-in each remaining cell by multiplying each respective row proportion by each respective column proportion gives…

Relative frequency table

	Grades	Popularity	Sports
Rural	0.223	0.130	0.092
Suburban	0.226	0.132	0.093
Urban	0.052	0.031	0.021

- Convert the relative frequencies back to counts by multiplying by the total count of individuals (335 in this case)

	Grades	Popularity	Sports
Rural	0.223	0.130	0.092
Suburban	0.226	0.132	0.093
Urban	0.052	0.031	0.021

x 335

	Grades	Popularity	Sports
Rural	75	43	31
Suburban	76	44	31
Urban	17	11	7

Performing the hypothesis test

- I can now compare the **expected counts** under H_0 true to the **observed counts.**

observed counts

	Grades	Popularity	Sports
Rural	57	50	42
Suburb	87	42	22
Urban	24	6	5

expected counts

	Grades	Popularity	Sports
Rural	75	43	31
Suburb	76	44	31
Urban	17	11	7

- For each cell (there are 9 in this case), we will compare the observed and expected counts to create a test statistic for our hypothesis test.

Performing the hypothesis test

- The Chi-Square Statistic:

$$\chi^2 = \text{sum of all values } \frac{(\textit{Observed - Expected})^2}{\textit{Expected}}$$

observed counts

	Grades	Popularity	Sports
Rural	57	50	42
Suburb	87	42	22
Urban	24	6	5

expected counts

	Grades	Popularity	Sports
Rural	75	43	31
Suburb	76	44	31
Urban	17	11	7

Performing the hypothesis test

- The Chi-Square Statistic:

$$\chi^2 = \frac{(57-75)^2}{75} + \frac{(50-43)^2}{43} + \ldots + \frac{(5-7)^2}{7} = 19.38$$

observed counts

	Grades	Popularity	Sports
Rural	57	50	42
Suburb	87	42	22
Urban	24	6	5

expected counts

	Grades	Popularity	Sports
Rural	75	43	31
Suburb	76	44	31
Urban	17	11	7

Performing the hypothesis test

- Making the decision: $\chi^2 = 19.38$
- Table 10.7 gives the critical values of χ^2 for two significance levels, 0.05 and 0.01.

Table 10.7 Critical Values of χ^2: Reject H_0 Only If $\chi^2 >$ Critical Value

Table size (rows × columns)	Significance level	
	0.05	0.01
2 × 2	3.841	6.635
2 × 3 or 3 × 2	5.991	9.210
3 × 3	9.488	13.277
2 × 4 or 4 × 2	7.815	11.345
2 × 5 or 5 × 2	9.488	13.277

Our test is significant at the 0.01 level because $\chi^2 = 19.38$ is larger than the 0.01 critical value of 13.277.

Performing the hypothesis test

- Conclusion: There is statistically significant evidence that the school type is related to which activity students find most important.