EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020
RAZ: Rebecca A. Zárate, MA

Overview

- Probability is Weird
- Contingency Tables
  - Marginal Probability
  - Joint Probability
  - Conditional Probability

Probability is Weird

- Sophie is 30 years old and majored in Engineering. As a student, she participated in protests advocating for equality and was deeply concerned with social justice. **Which is more probable?**

A) Sophie is a social media influencer.
B) Sophie is a social media influencer *and* is active in the feminist movement.

Probability is Weird

- As a student, she participated in protests advocating for equality and was deeply concerned with social justice. Which is more probable?

A) Sophie is a social media influencer.

B) Sophie is a social media influencer *and* is active in the feminist movement.

If you think about it, the probability of being a social media influencer might be .10 but we also have to factor in the probability of being in the feminist movement (maybe that's .20). That is a subset of all influencers, so the probability is lower to be a social media influencer AND active in the feminist movement. To determine the probability of BOTH we need to multiply them…

$$.10 \times .20 = .02$$

The probability of being both (.02) is much lower than the probability of just being a social media influencer (.10).

Contingency Tables

Contingency Tables

- A contingency table is a table that has two categorical variables represented.
- Each cell corresponds to the frequency (i.e. count, tally) of seeing certain levels of the two categorical variables.
  - For example, we have a table with Tests Grades and whether or not a student studied.

	Test Grade			
	A	B	C	D
Studied	11	17	7	3
Didn't Study	1	4	12	15

Contingency Tables

	Test Grade			
	A	B	C	D
Studied	11	17	7	3
Didn't Study	1	4	12	15

What kind of variables are we working with?

Contingency Tables

	Test Grade			
	A	B	C	D
Studied	11	17	7	3
Didn't Study	1	4	12	15

Here we have some frequencies for two categorical variables 1) Studied vs. Didn't Study and 2) the letter grade)

Marginal Probability

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

What percent of the students studied?
What percent of the students received a B?

Marginal Probability

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

What percent of the students studied?

38/70 = .54 or 54%

What proportion of the students received a B?
21/70 = .30
or 30%

These are called "Marginal Probabilities": The probability of seeing someone of a certain category (ex. Study vs. No Study) regardless of the other variable (the grade they got) or vise versa.

Joint Probability

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

What is the probability I select someone at random that Didn't Study AND got an A?

Joint Probability

What is the probability I select someone at random that Didn't Study **and** got an A?

1.4%

1/70 = .014

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

This is called "Joint Probability", the probability of two things happening, ex. getting an A and not studying.

Wording Matters with Conditionals

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

Of those that got a D, what percent studied?

Of those that did not study, what percent got a D?

How is this different from before?

Conditional Probability

$$\text{Conditional Probability} = \frac{\text{Joint Frequency}}{\text{Marginal Total}}$$

Of those that got a D, what percent studied?

Now the thing we divide by is restricted to a certain condition, here those that got a D.

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

Wording Matters

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38
Didn't Study	1	4	12	15	32
Total	12	21	19	18	70

Of those that got a D, what percent Studied?

3/18 = .16, 16%

Of those that Didn't Study, what percent got a D?

15/32 = .47, 47%

- They sound like the same question, but they are not…
  - Helpful tip, identify your denominator first
    - Ex. "Of (certain condition)" is the denominator

Marginal Probabilities

	Test Grade				Total
	A	B	C	D	
Studied	11	17	7	3	38 (.54)
Didn't Study	1	4	12	15	32 (.46)
Total	**12 (.17)**	21 (.30)	19 (.27)	18 (.26)	70

- These are the probabilities of seeing just one condition, ex. the probability of someone who got an "A" 12/70 = 0.17.

Joint Probabilities

	Test Grade				Total
	A	B	C	D	
Studied	**11 (.15)**	17 (.24)	7 (.10)	3 (.04)	38
Didn't Study	1 (.01)	4 (.05)	12 (.17)	15 (.21)	32
Total	12	21	19	18	70

- These are the probabilities of seeing both conditions, ex. the probability of someone who got an "A" AND studied is $11/70 = 0.15$.

Conditional Probability (Studied-Didn't Study)

	Test Grade				Total
	A	B	C	D	
Studied	11/38 = .29	17/38 = .45	7/38 = .18	3/38 = .08	**38**
Didn't Study	1/32 = .03	4/32 = .13	12/32 = .38	15/32 = .47	**32**
Total	12	21	19	18	70

Denominators

- Of those that Studied ("Given someone Studied") the probability someone got an A is .29.
  - Here the denominator is the ROW sum

Conditional Probability (Test Grade)

	Test Grade				Total
	A	B	C	D	
Studied	11/12 = .91	17/21 = .81	7/19 = .36	3/18 = .17	38
Didn't Study	1/12 = .08	4/21 = .19	12/19 = .63	15/18 = .83	32
Total	**12**	**21**	**19**	**18**	**70**

Denominators

- Of those that got an A ("Given someone got an A") the probability someone Studied is .91.
  - Here the denominator is the COLUMN sum

Contingency Tables Recap

- Marginal Probability
  - The probability of seeing one certain thing
- Joint Probability
  - The probability of seeing two certain things
- Conditional Probability
  - The probability of seeing one certain thing, GIVEN you already saw the other certain thing

Up Next…

- Contingency tables and their associated probabilities are descriptive in nature, but what if we want to statistically test if two categorical variables are independent from each other, ex. Is your test grade "independent" of whether you Studied or Did Not Study? To test this we need…

Chi-Squared Test of Independence