MOD 11: TWO-WAY TABLES WITH AN INTRODUCTION TO PROBABILITY

Learning Goals

- Use a two-way table to analyze the association between two categorical variables.
- Given a two-way table, interpret (in context) joint, marginal, and conditional probability.
- Create a hypothetical two-way table to answer complex probability questions.

1) What is the difference between categorical data and quantitative data? Give examples of each.

2) Use the data from our class survey to fill in the table.

		Eye Color				
		Blue	Green	Brown	Black	**Total**
Hair Color	Blonde					
	Red					
	Brown					
	Black					
	Total					

a) How many students in our class have brown hair and green eyes?

b) How many have black hair and brown eyes?

c) How many students in our class have brown hair?

d) How many students in our class have brown eyes?

e) What do the values in the bottom row represent (excluding the lower right-hand corner)?

f) What do the values in the right hand column represent (excluding the lower right-hand corner)?

By the way … for any two-way table representing two categorical variables the totals in the bottom row and right-hand column (excluding the lower right-hand corner) are the ***marginal distributions****.*

g) What does the value in the bottom right-hand corner represent? How did you get this number? Did you add the values in the right-hand column, or did you add the values along the bottom row? Would either method always work? Why or why not?

h) What fraction of students in our class have brown hair and brown eyes? What fraction of students in our class have green eyes?

i) Suppose there is a statistics class at Cuyamaca with 40 students and 20 of them earn an A on the first exam (the rest earned a C). Further suppose there is a statistics class at UCSD with 500 students and 20 of them earn an A on the first exam (again, the rest earned a C). Since 20 students in each class earned an A on the first exam, can we assume the two classes performed the same on that exam? Why or why not? How could you make the difference evident even though the number of students earning an A is the same in both classes?

3) A few years ago, Myra Snell (the inspiration behind Cuyamaca's "Stats Academy") gave her PreStats students a quiz containing a question from a national statistics exam. Twenty-nine students took the quiz. She compared their performance on a particular question to the performance of the 1470 students included in the national sample. 24 of her students got the item right compared to 1001 of the national sample. Did Myra's PreStats students do as well as the national sample? Use math to support your conclusion.

4) This data comes from a study of the factors that impact birth weight. Here the variable *Visit Doctor* indicates whether a woman visited a physician during the first trimester of

her pregnancy. The variable *Low Weight* indicates whether a baby was born weighing under 2500 grams.

Low Weight

Visit Doctor	No	Yes	Row Totals
Yes	66	23	
No	64	36	
Column Totals			

Question to be investigated: *Does visiting a doctor during the early stages of pregnancy seem to be associated with a lower incidence of low weight births?*

a) Complete the table to find the marginal distributions, and then fill in the bottom right-hand corner.

b) Identify the **explanatory** and **response** variables.

c) Let's set up the fractions we will use to make the comparison and draw our conclusions. To answer a research question such as the one we are investigating, we always compare the categories of the explanatory variable, so the totals for these categories will be the bottoms of our fractions. The tops of our fractions come from the one category of the response variable asked about in the research question. Set up the fractions and compute the percentages you will use to make the comparison and draw your conclusions.

d) What is the percent decrease in low-weight births for women who visit the doctor regularly in the early stages of pregnancy?

e) Draw conclusions to answer the question, *Does visiting a doctor during the early stages of pregnancy seem to be associated with a lower incidence of low weight births?*

5) This table is based on records of accidents compiled by a State Highway Safety and Motor Vehicles Office (the marginal distributions and the lower right-hand corner have been filled in for you). Are people less likely to have a fatal accident if they are wearing a seatbelt? Be sure to clearly identify the explanatory and response variables and use either a "percent increase" or "percent decrease" argument to support your conclusion.

		Injury		
Seat Belt		**Nonfatal Injury**	**Fatal Injury**	**Row Total**
	Seat belt	412,368	510	412,878
	No seat belt	162,527	1,601	164,128
	Column Total	574,895	2,111	577,006

6) Four hundred seventy-eight students in grades 4 through 6 in selected schools in Michigan, were asked the following question.

Which of the following would make you popular among your friends? Rank in order.

- *Making good grades*
- *Having lots of money*
- *Being good at sports*
- *Being handsome or pretty*

The table below lists the number of students (by gender) who gave the indicated factor a ranking of 1 (most important factor in making them popular amongst their friends).

Most Important Popularity Factor

Gender		Grades	Money	Sports	Looks	**Row Totals**
	Girl	55	17	38	141	
	Boy	39	17	127	44	
	Column Totals					

a) What proportion of the total number of students considers looks to be the most important factor in making them popular amongst their friends?

b) If we were to randomly select a boy, what is the probability that he would think Money is the most important factor in making them popular amongst their friends?

To calculate the proportion in number 6a, we used two numbers in the margin that relate to just one of the categorical variables (Student Response). This is therefore called a ***marginal proportion****. It's actually a* ***marginal probability*** *as well (same thing).*

However, in part b) we did not calculate a marginal probability or marginal proportion. Instead we used a ***conditional proportion*** (a proportion that is "conditioned" on the explanatory variable) to find the conditional probability. We also used conditional proportions in numbers 4 and 5 above (only we didn't know we were calculating conditional proportions when we were working on those problems).

7) **Popularity Revisited:** Four hundred seventy-eight students in grades 4 through 6 in selected schools in Michigan, were asked the following question.

Which of the following would make you popular among your friends? Rank in order.

- *Making good grades*
- *Having lots of money*
- *Being good at sports*
- *Being handsome or pretty*

The table below lists the number of students (by gender) who gave the indicated factor a ranking of 1 (most important factor in making them popular amongst their friends).

		Most Important Popularity Factor				
		Grades	Money	Sports	Looks	**Row Totals**
Gender	Girl	55	17	38	141	
	Boy	39	17	127	44	
	Column Totals					

a) Suppose we randomly select a student. What is the probability that the student is a boy? Is this a marginal or conditional probability? How do you know?

b) What is the probability that a student ranked grades as the most important factor in making him/her popular amongst his/her friends? Write your response in a sentence and then write it using the appropriate probability notation (if you don't know how to do this, be sure to ask). Is this a marginal or conditional probability? How do you know? **Show the instructor your answers to part b**.

c) What is the probability that a boy ranked sports as the most important factor in making him popular amongst his friends? Is this a marginal or conditional probability? How do you know? Write your response in a sentence and then write it using the appropriate probability notation. **Show the instructor your answers to part c**.

d) What is the probability that a girl ranked looks as the most important factor in making her popular amongst her friends? Is this a marginal or conditional probability? How do you know? Write your response in a sentence and then write it using the appropriate probability notation. **Show the instructor your answers to part d**.

e) **Joint Probability**: Suppose we randomly select a student. What is the probability that the student is a boy **and** the student ranked grades as the most important factor in making him popular amongst his friends? Write your response in a sentence and then write it using the appropriate probability notation. **Show the instructor your answers to part e**.

f) Find and **interpret** the *following* probabilities and then indicate whether you have found a marginal probability, a conditional probability, or a joint probability.

P(Money|Girl)

P(Sports)

P(Girl and Sports)

8) Many people who visit clinics to be tested for HIV (the virus that causes AIDS) do not return to learn the test results. Consequently, clinics now use rapid HIV tests that give the result while the client waits.

The trade-off for fast results is that rapid tests are less accurate than slower laboratory tests – especially when the test is applied within four to six weeks after exposure.

Using a rapid test within four to six weeks of exposure, we can assume that 95% of HIV infections are detected (test positive).

Using a rapid test within four to six weeks of exposure, we can assume there is 0.4% chance that the rapid test will be positive for an uninfected person.

In Malawi in 2012, the HIV infection rate was 10,800 out of every 100,000 people.

	Infected	**Not Infected**	**Row Totals**
Test HIV Positive			
Test HIV Negative			
Column Totals			100,000

What is the probability of a false positive test result? A false positive is the probability that a person is not infected given the test is positive. Show your work.

MOD 11 – LET'S PRACTICE

1) Here is the distribution of conferred doctorates by field and gender in 2009. NOTE: ALL FIELDS ARE DOCTORATES.

	Engineering	Physical Sciences	Life Sciences	Social Sciences	Education	Other	Row Totals
Male	6006	5868	5180	3259	2160	3865	26,338
Female	1623	2450	6212	4575	4370	3960	23,190
Column Totals	7629	8318	11,392	7834	6530	7825	49,528

Use the table to find and interpret each of the following probabilities.

a) P(Female and Education)

b) P(Life Sciences | Male)

c) P(Female | Life Sciences)

d) P(Engineering)

2) Polygraph lie-detector machines are commonly used in criminal investigations. The device measures nervous excitement, operating on the idea that if a person is telling the truth they will remain calm.

- The American Polygraph Association claims that polygraphs accurately identify liars 90% of the time. In other words, if a person lied the test results will be positive 90% of the time.
- Although the polygraph test is good at identifying liars, there is a 50% chance that the polygraph test will say a truthful person is lying.

A police force wants to predict the accuracy of lie-detector tests for 1,000 suspected criminals. Assume that 700 per 1,000 suspected criminals tell the truth during polygraph tests.

	Lie	**Truth**	**Row Totals**
Test result: negative			
Test result: positive			
Column Totals			1000

What is the probability that a person is telling the truth given that the lie detector test has a positive result?