

## ***Section 2.1: Categorical Variables***

### ***Objectives***

- 1) One categorical variable:
    - a. Frequency and relative frequency tables
    - b. Find proportions
    - c. What graphs are used to graph a categorical variable? Construct them
      - i. Bar graphs
      - ii. Pie graphs
  - 2) Two categorical variables:
    - a. Construct a two-way table
    - b. Find proportions
    - c. What graphs should be used? Construct them
      - i. Segmented bar graphs/charts
      - ii. Side by side bar graphs/charts
- 

***Example: Let's make up some data to introduce all these concepts.***

- a) *How do you commute to school? Use three categories. Assume we surveyed 100 students.*
- b) *Add the variable gender.*
- c) *Ask appropriate questions and answer them.*

## ***Section 2.1: Categorical Variables***

## Section 2.1: Categorical Variables

Read section 2.1, starting with page 48; then, answer the questions. Listen to the videos for this section in WILEYPLUS

- 1) In the book, read the description of DATA 2.1 and the Example 2.1.
  - What method of sampling was used?
  - The sample is
  - The population is
  - Is it a biased sample?
  
- 2) In the book, read example 2.2
  - The cases are
  - The variables are
  - Classify each as categorical or quantitative
  
- 3) Show a table listing possible responses for two cases

### Read the book: One categorical variable

- 4) Write the formula to find proportions.
  
  
  
  
  
  
  
  
  
  
- 5) Proportions are also called \_\_\_\_\_
  
  
- 6) Tables 2.1 and 2.2 display part of the DATASET 2.1. **answers to the survey “There is one true love for each person”** - Notice the difference between frequency tables (2.1) and relative frequency tables (2.2). Summarize them here:

Response.....	Frequency..... Count (how many in Each category)	Relative Frequency..... Proportion Percent as a decimal
Total		

- 7) Read example 2.3. Show how each of the relative frequencies on the table above were calculated.
  
  
  
  
  
  
  
  
  
  
- 8) What is the difference between a frequency table and a relative frequency table?

## Section 2.1: One Categorical Variables

### 9) Example 1: Talking About Sports

A survey in November 2012 asked a random sample of 2,000 US adults “How often do you talk about sports with family and friends?” The results are given in the following frequency table.

Response	Frequency	Relative Frequency - Proportion
Every day or nearly every day	302	
About once a week	277	
Occasionally	526	
Rarely or never	895	
TOTAL	2000	

- Complete the proportions column; show all calculations and round to three decimal places:
- What proportion rarely or never talk about sports?
- What percent of people in the sample talk about sports once a week or more?

### Visualizing the data in One Categorical Variable:

- 10) What **graphs** are used to display information about categorical variables?

Example: Males and Female students in our class – create a frequency and relative frequency table and use appropriate graphs for the data.

- 11) The vertical axis on a bar graph represents \_\_\_\_\_ or \_\_\_\_\_.

- 12) **Notation** for proportions

In the Fall 2017, 59.3% of students in Takoma Park were female while for the whole college, 53.4% were female. Write as proportions with correct notation

	Notation
sample	
population	

## Section 2.1: Two Categorical Variables

**Read the book: Two categorical variables** – read all examples from the book.

- 13) To show the relationship of two categorical variables we use a \_\_\_\_\_.
- 14) In the book Example: “Is there one true love for each person?” Look at the two way table 2.4 (on page 52) and read example 2.5

### TO DO IN CLASS

**15) Example 2: Relationship Status and Gender** (similar to example 2.5 and 2.7)

169 college students were asked about relationship status and gender. The results are given in the following two-way table.

	Female	Male	Total
In a relationship	32	10	42
It's complicated	12	7	19
Single	63	45	108
Total	107	62	169

Show all calculations – round all proportions to three decimal places:

- a). What proportion of students in this sample are in a relationship?
- b). What proportion of females in this sample are in a relationship?
- c). What proportion of the people who are in a relationship in this sample are female?
- d). What proportion of males in this sample are in a relationship?
- e). Using  $\hat{p}_F$  to represent the proportion of females in a relationship and  $\hat{p}_M$  to represent the proportion of males in a relationship, find the difference in proportions  $\hat{p}_F - \hat{p}_M$ .
- f) Does it appear to be an association between gender and being in a relationship?

## **Section 2.1: Two Categorical Variables**

### **16) Example 3: Handedness and Occupation** (similar to example 2.6)

In a study of handedness in occupations, 10 out of 118 psychiatrists were left-handed, 26 out of 148 architects were left-handed, 5 of 132 orthopedic surgeons were left-handed, and 16 of 105 lawyers were left-handed.

a). Make a two-way table of this relationship.

b). What proportion of all the people in the sample are left-handed? Show calculations

c). What proportion of architects are right handed? Show calculations

17) Read the section:

**Visualizing a Relationship between Two Categorical Variables** (pgs. 54, 55) - Segmented and side by side bar charts

## Section 2.1: Two Categorical Variables

### 18) Book problem: 2.22 - Can Dogs Smell Cancer?

Scientists are working to train dogs to smell cancer, including early stage cancer that might not be detected with other means. In previous studies, dogs have been able to distinguish the smell of bladder cancer, lung cancer, and breast cancer. Now, it appears that a dog in Japan has been trained to smell bowel cancer.<sup>15</sup> “Dog Detects Bowel Cancer,” *CNN Health Online*, January 31, 2011. Researchers collected breath and stool samples from patients with bowel cancer as well as from healthy people. The dog was given five samples in each test, one from a patient with cancer and four from healthy volunteers. The dog correctly selected the cancer sample in 33 out of 36 breath tests and in 37 out of 38 stool tests.

- (a) The cases in this study are the individual tests. What are the variables?
- (b) Make a two-way table displaying the results of the study. Include the totals.
- (c) What proportion of the breath samples did the dog get correct? What proportion of the stool samples did the dog get correct?
- (d) Of all the tests the dog got correct, what proportion were stool tests?

## Statkey and Categorical Variables

### 19) Working with STATKEY – to access STATKEY try one of the following options:

1. Log into WILEYPLUS, select one section of the book (any is ok). Click on EXPLORE, click on STATKEY
2. Or go to [Lock5stat.com/statkey](http://Lock5stat.com/statkey)
3. My website has also a statkey link

### 20) Get into STATKEY

- On the **Descriptive Statistics and Graphs** column
- Click on **One Categorical Variable**
- On the first drop down menu, select one of the datasets
- Pay attention to the tables and graphs displayed and think what questions you can answer based on the data

### 21) Click on STATKEY again

- Click on **Two Categorical Variables**
- Select a dataset
- Pay attention to the tables and graphs displayed and think what questions you can answer based on the data

## Section 2.1: Two Categorical Variables

### 22) Painkillers and Miscarriage

A recent study<sup>15</sup> examined the link between miscarriage and the use of painkillers during pregnancy. Scientists interviewed 1009 women soon after they got positive results from pregnancy tests about their use of painkillers around the time of conception or in the early weeks of pregnancy. The researchers then recorded which of the pregnancies were successfully carried to term. The results are in Table 2.10.

23) Table 2.10

*Does the use of painkillers increase the risk of miscarriage?*

	Miscarriage	Total
Aspirin	5	22
Ibuprofen	13	53
Acetaminophen	24	172
No painkiller	103	762
Total	145	1009

- (a) What percent of the pregnancies ended in miscarriage?
- (b) Compute the percent of miscarriages for each of the four groups. Discuss the results. Add a column to the table shown above with this percentages (to one decimal place)
- (c) Is this an experiment or an observational study? Describe how confounding variables might affect the results.
- (d) Aspirin and ibuprofen belong to a class of medications called nonsteroidal anti-inflammatory drugs, or NSAIDs. What percent of women taking NSAIDs miscarried? Does the use of NSAIDs appear to increase the risk of miscarrying? Does the use of acetaminophen appear to increase the risk? What advice would you give pregnant women?
- (e) Is Table 2.10 a two-way table? If not, construct one for these data.
- (f) What percent of all women who miscarried had taken no painkillers?



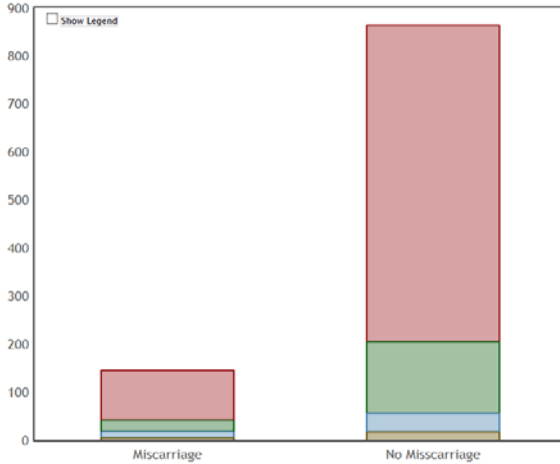
## Section 2.1: Two Categorical Variables

### Painkillers and Miscarriage

Row Proportions – How did we find each proportion? Write a possible question for a few of them.

#### StatKey Descriptive Statistics for Two Categorical Variables

Painkillers and Miscarriage Edit Data Upload File Change Column(s)



#### Counts Table Switch Variables

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	5	17	22
Ibuprofen	13	40	53
Acetaminophen	24	148	172
No Painkiller	103	659	762
Total	145	864	1009

#### Proportions Row Column Overall

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	0.227	0.773	1
Ibuprofen	0.245	0.755	1
Acetaminophen	0.14	0.86	1
No Painkiller	0.135	0.865	1
Total	0.144	0.856	1

## Section 2.1: Two Categorical Variables

### Painkillers and Miscarriage

Column Proportions – How did we find each proportion? Write a possible question for a few of them.

#### Counts Table Switch Variables

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	5	17	22
Ibuprofen	13	40	53
Acetaminophen	24	148	172
No Painkiller	103	659	762
Total	145	864	1009

#### Proportions Row Column Overall

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	0.034	0.02	0.022
Ibuprofen	0.09	0.046	0.053
Acetaminophen	0.166	0.171	0.17
No Painkiller	0.71	0.763	0.755
Total	1	1	1

## Section 2.1: Two Categorical Variables

### Painkillers and Miscarriage

Overall Proportions – How did we find each proportion? Write a possible question for a few of them.

#### Counts Table [Switch Variables](#)

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	5	17	22
Ibuprofen	13	40	53
Acetaminophen	24	148	172
No Painkiller	103	659	762
Total	145	864	1009

#### Proportions [Row](#) [Column](#) [Overall](#)

undefined \ undefined	Miscarriage	No Miscarriage	Total
Aspirin	0.005	0.017	0.022
Ibuprofen	0.013	0.04	0.053
Acetaminophen	0.024	0.147	0.17
No Painkiller	0.102	0.653	0.755
Total	0.144	0.856	1