

STAB22 TUT21

Chong Chen
University of Toronto, Scarborough
Department of Computer and Mathematical Sciences

January 25, 2019

1 Recall

Categorical variable: a variable that places an individual into one of two or more groups or categories.

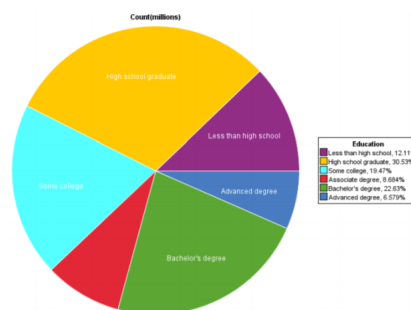
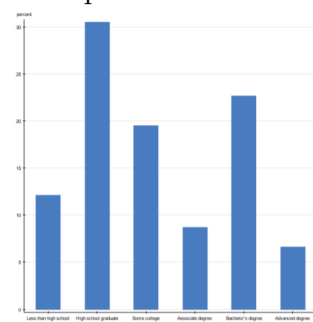
2 Displaying Distributions with Graphs

2.1

The distribution of a categorical variable lists the categories and gives either the count or percent of individuals who fall in each category.

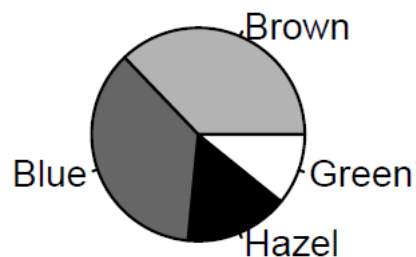
There are two ways to display one categorical data:

1. A bar chart
2. A pie chart



2.2 Example

The pie chart below shows eye color in 592 statistics students.



Approximately how many of these statistics students have brown eyes?

- (a) 33
- (b) 50
- (c) 150
- (d) 200

3 Conditional distribution

3.1

When we condition on the value of one variable and calculate the distribution of the other variable, we obtain a conditional distribution.

3.2 Examples

3.2.1

For patients who have coronary heart disease, some people believe that owning a pet helps in their survival. One study examined the relationship between survival after one year and pet ownership. The data are shown below:

After one year	No	Yes
Alive	18	60
Dead	11	8

- (a) Out of the patients who owned a pet, what percentage were alive after one year?
- (b) Out of the patients who did not own a pet, what percentage were alive after one year?
- (c) Does the data suggest that owning a pet is associated with survival? Explain briefly

3.2.2

The contingency table below shows the favourite leisure activity of 50 adults - 20 men and 30 women.

	Dance	Sports	TV
Men	2	10	8
Women	16	6	8

What percent of the 50 adults have Dance as favourite leisure activity?

- (a) 9
- (b) 18
- (c) 33
- (d) 36

3.2.3

The contingency table below shows the favourite leisure activity of 50 adults - 20 men and 30 women.

	Dance	Sports	TV
Men	2	10	8
Women	16	6	8

Which of the following statements about these 50 adults is/are correct?

- (I) Men are more likely to have Dance as their favourite leisure activity than women.
- (II) Men and women are equally likely to have TV as their favourite leisure activity.
- (III) Women are twice as likely to have Sports as their favourite

leisure activity than men are to have Dance as their favourite leisure activity.

- (a) Only statement (I) is correct
- (b) Only statement (II) is correct
- (c) Only statement (III) is correct
- (d) Exactly two statements are correct
- (e) The three statements are correct

4 Two-way tables

For two categorical variables the raw data are summarized in a two-way table that gives counts of observations for each combination of values of the variables.

Example:

A survey of 17,096 students in U.S. four-year colleges collected information on drinking behaviour and alcohol-related problems. The researchers defined “frequent binge drinking” as having five and more drinks in a row three or more times in the past two weeks. Here is the two-way table classifying students by gender and whether or not they are frequent binge drinkers:

Frequent binge drinker:	Men	Women
Yes	1630	1684
No	5550	8232

Gender is a column variable.

Binge drinking is a row variable.

Combinations of values for two variables are called cells.