Graphical Descriptive Statistics
Part I: Depicting a single variable

Eco 2470: Economic Statistics

Fall, 2019. Chaoyi Chen

(Chapters 2-3)

Interval Data

- Also known as quantitative or numeric data
- Numbers that have meaning
- Examples: age, years schooling, wage, GDP, foul shot percentage, exchange rate

Bank of Canada, 2011 Yearly Average Exchange Rates

Country	Currency	value in Canadian Currency
China	renminbi	0.1531
Europe	Euro	1.3767
Japan	yen	0.01242
United States	dollar	0.9890692

Ordinal Data

- Numbers denote ordered categories

- Only the <u>order</u> matters

- E.g. Highest Degree Completed

  1 (none), 2 (elementary), 3 (high school), 4 (university)

- E.g. Course Evaluations:

  1 (poor), 2 (fair), 3 (good), 4 (very good), 5 (excellent)

Comparing Interval & Ordinal Data

ATP Rankings Tennis Rankings (Men's Singles), Aug 20, 2012

Name	Rank	Points
Roger Federer	1	12,165
Novak Djokovic	2	11,270
Rafael Nadal	3	8,715
Andy Murray	4	7,290

Nominal Data

- Also known as Categorical or Qualitative
- Numeric values just denote a name or category
- Have no mean as a number
- Examples:
  - Sex: 0 (male), 1 (female)
  - Region: 1 (Ontario), 2 (Quebec), 3 (PEI), …
  - Postal Code

Frequency Definitions

- <u>Frequency</u>: Number of observations falling into a group or category

- <u>Relative Frequency</u>: Proportion of observations falling into a group or category

- <u>Cumulative Relative Frequency</u>: Proportion of observations falling into a group and all previous groups
  - Applies only to ordered groups
  - Applies to Ordinal, but not nominal data

Example: Health, malnutrition and obesity among married women in India

Sources:Demographic and Health Surveys (DHS), Sarah Salih's U. of Guelph MA paper.
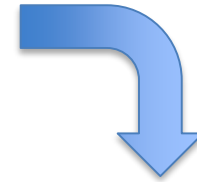
BMI=Body Mass Index = weigh/(height^2)

Health Category	Frequency	Relative Frequency	Cumulative Relative Frequency
Underweight (BMI<18.5)	23,522	0.2896 (= 23,522/103,288)	0.2896
Normal	51,528	0.4589 (=51,528/103,288)	0.7485 (=0.2896+0.4589
Overweight (BMI>23)	28,238	0.2515 (=28,238/103,288)	1 (=0.7485+0.2515)
Total	103,288	1	N/A

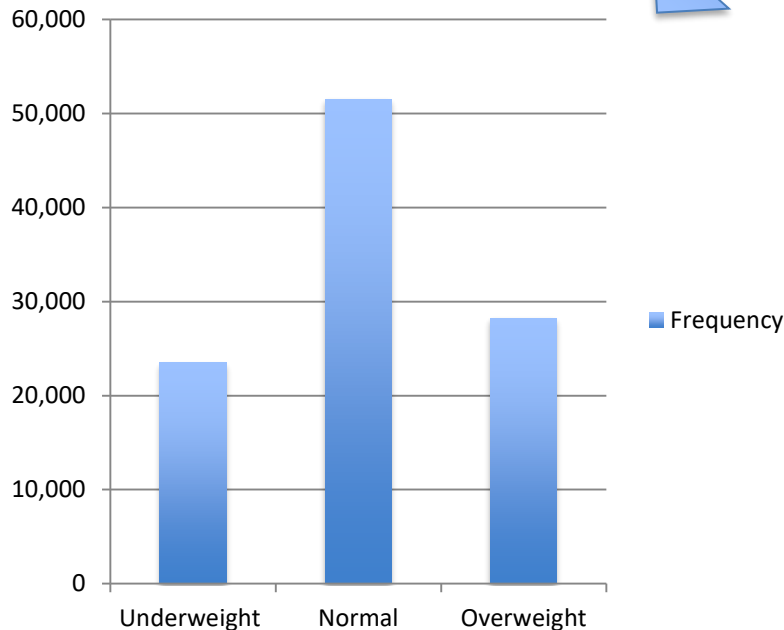# Bar and Pie Charts

- <u>Bar Chart</u>: Uses bar heights to display <u>frequencies</u> for each group.

- <u>Relative Frequency Bar Chart</u>: Same, except that is displays <u>relative frequencies</u>

- <u>Pie Chart</u>: Uses the size of "pizza slices" to display <u>relative frequencies</u>

- <u>Hint</u>: Make your frequency/relative frequency table first. Then use it to make your charts.

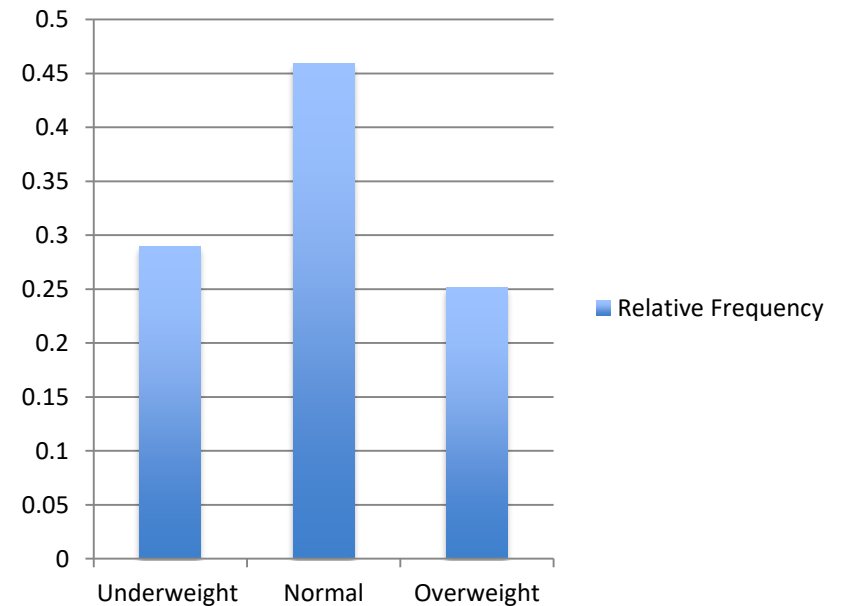# Health Status of Married Women in India (con't)

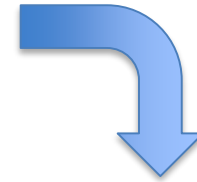| Health Category	Frequency	Relative Frequency
Underweight	23,522	0.29
Normal	51,528	0.459
Overweight	28,238	0.252

Bar Chart (Frequency)

Relative Frequency Bar Chart

Health Status of Married Women in India (con't)

Health Category	Frequency	Relative Frequency
Underweight	23,522	0.29
Normal	51,528	0.459
Overweight	28,238	0.252

Bar Chart (Frequency)

Relative Frequency Bar Chart
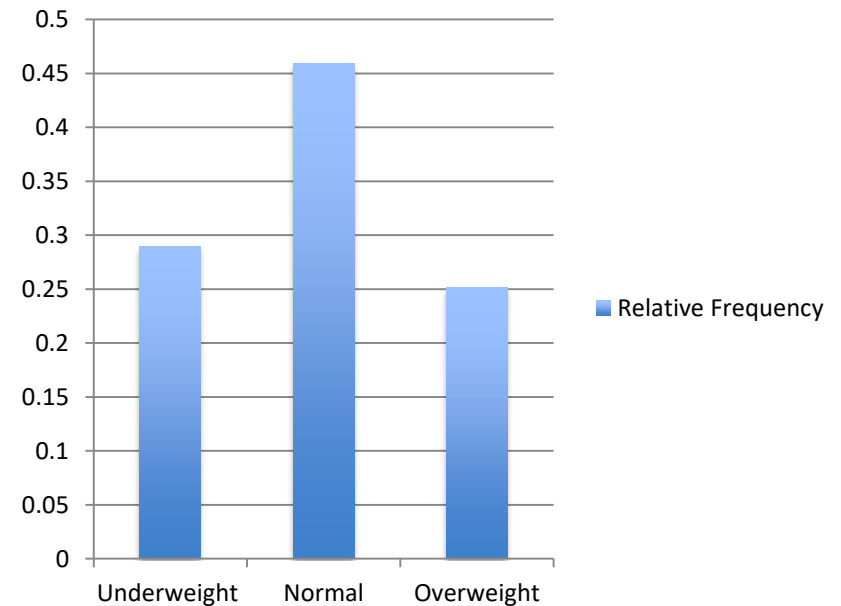
Health Status of Married Women in India (con't)

Health Category	Frequency	Relative Frequency
Underweight	23,522	0.29
Normal	51,528	0.459
Overweight	28,238	0.252

Pie Chart (Relative Frequency)

Pie chart showing: 0.2896 Underweight, 0.4589 Normal, 0.2515 Overweight
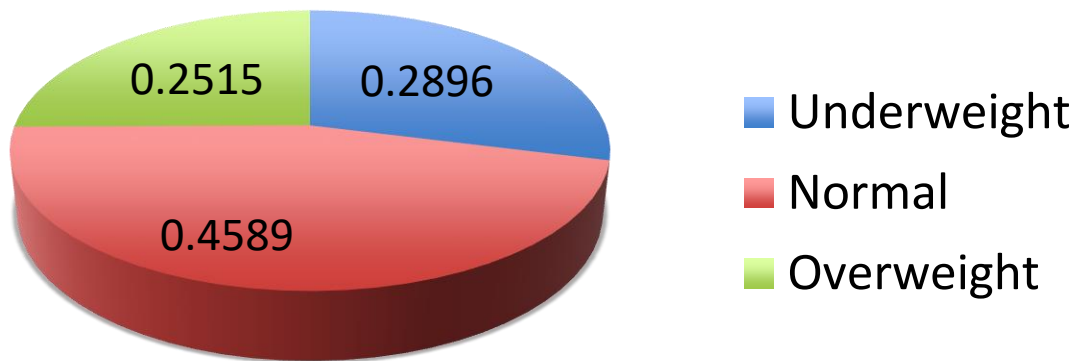
Histograms

- Nominal and Ordinal data: pre-defined groups
- Easy to calculate frequencies & plot bar chart
- Interval Data: No predefined groups
- What to do?
- Solution: Define are own groups,
- Except that they are no longer called groups
- They are now called <u>classes</u>
- Then calculate frequencies and make a bar chart
- Except it is no longer called a bar chart
- It is called a <u>histogram</u>

Example: US income distribution

- Survey of 114,384 households by U.S. Census Bureau.

- Record 2006 household income in US dollars

- Interval data since this is a meaningful dollar quantity

- Make classes based on income intervals:

Class 1 (0-10k), class 2 (10k-20k), …

- Calculate frequency for each class

US income distribution (con't)

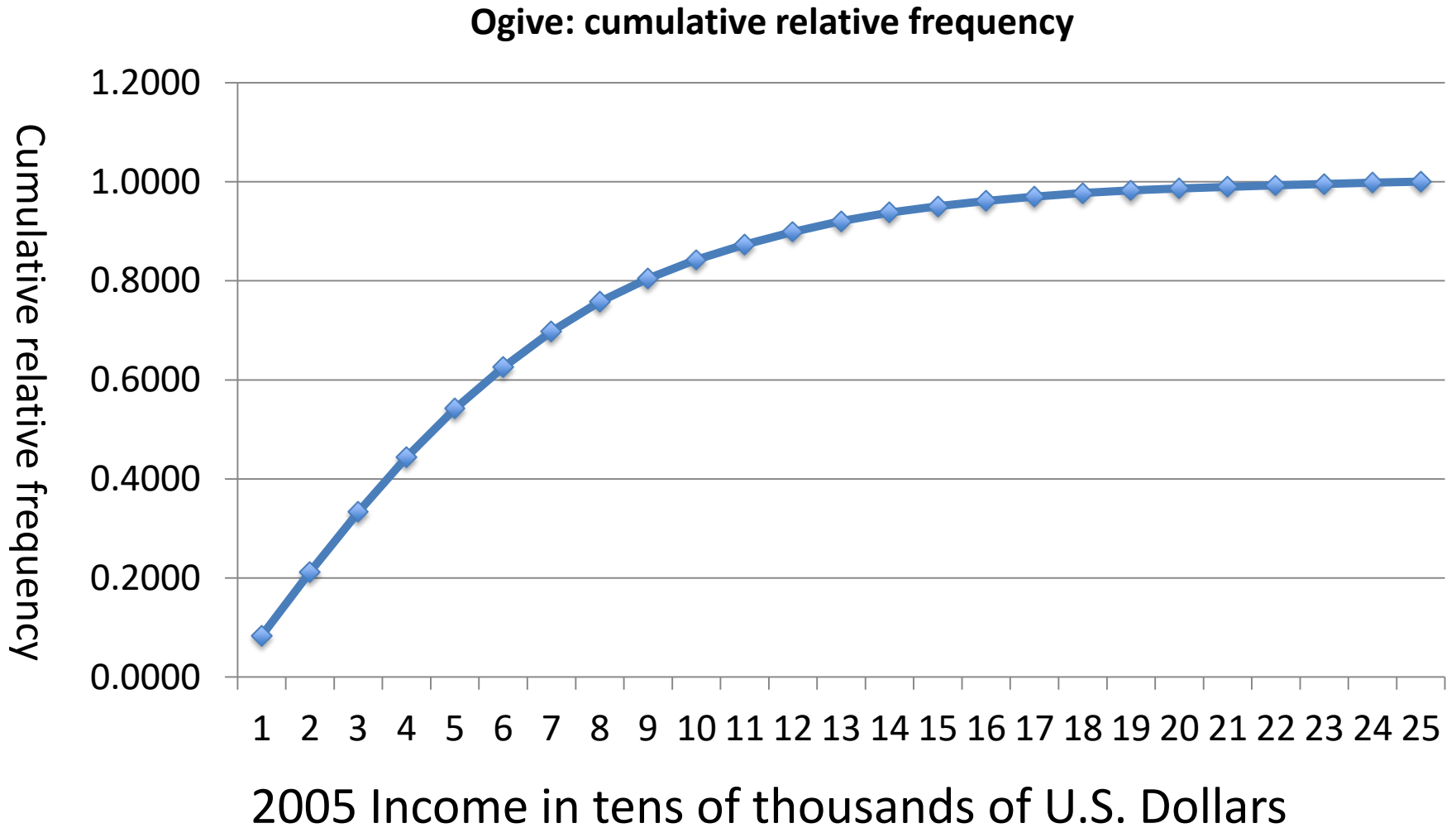| Class	Income more than	and less than	frequency (no. of households
1	0	10,000	9,401
2	10,000	20,000	14,447
3	20,000	30,000	13,642
4	30,000	40,000	12,388
5	40,000	50,000	11,028
6	50,000	60,000	9,352

Etc.

Cumulative Relative Frequency of Income

Class	Income more than	and less than	frequency (no. of households)	Total no. of households	relative frequency	cumulative relative frequence
1	0	10,000	9,401	112,364	0.0837	**0.0837**
2	10,000	20,000	14,447	112,364	0.1286	**0.2122**
3	20,000	30,000	13,642	112,364	0.1214	**0.3336**
4	30,000	40,000	12,388	112,364	0.1102	**0.4439**
5	40,000	50,000	11,028	112,364	0.0981	**0.5420**
6	50,000	60,000	9,352	112,364	0.0832	**0.6253**

Interpretation: Approximately 60 percent of households had income under 50,000.

Ogive (Graph of cumulative relative frequency)

Ogive: cumulative relative frequency

Interpreting the Ogive

Ogive: cumulative relative frequency

54 percent of households have
Income less than $50 thousand.

92 percent of households
Have income less than
$130 thousand

2005 Income in tens of thousands of U.S. Dollars
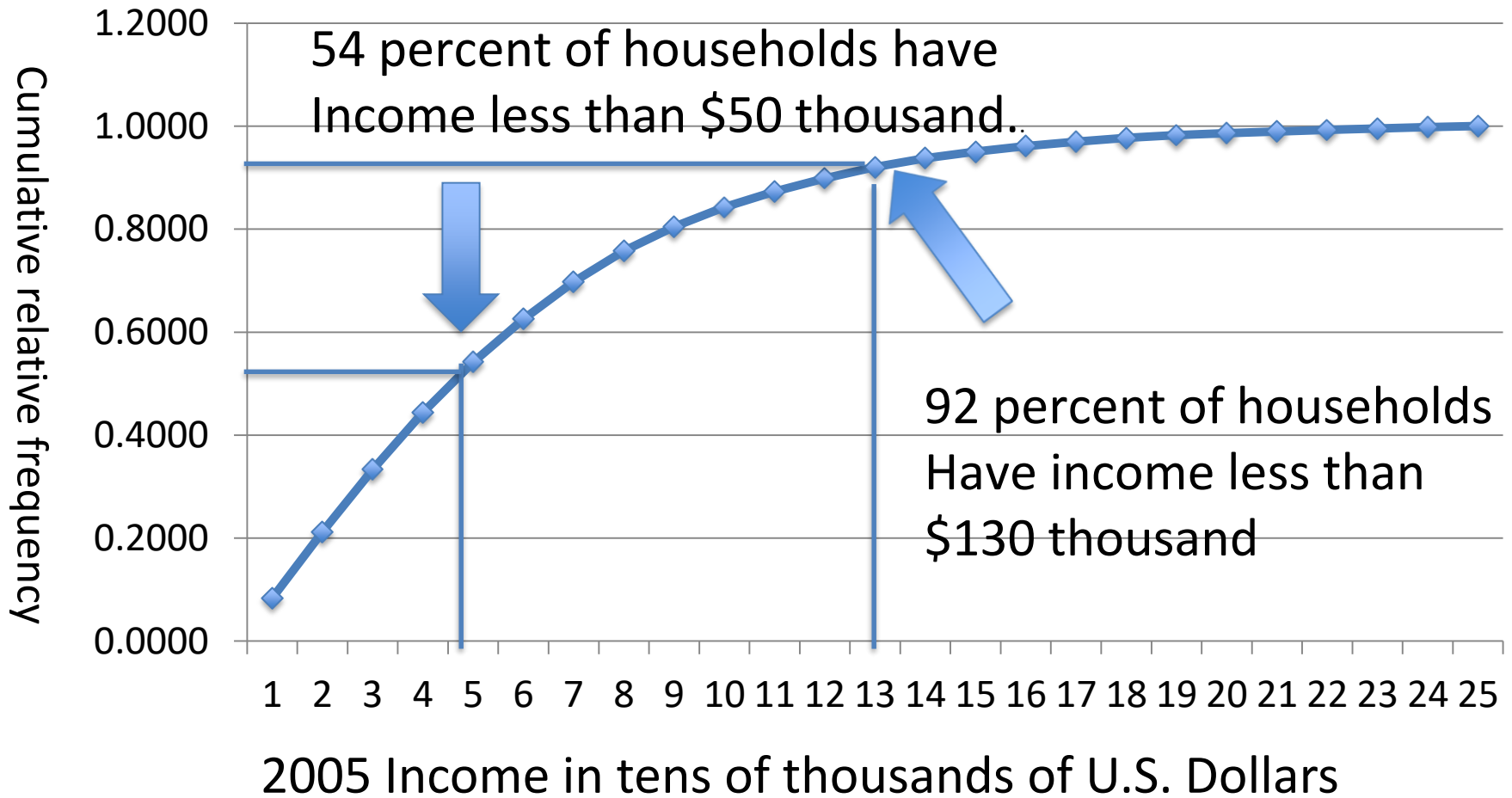
How to select number and width of classes for histogram and ogive?

Approximate Number of Classes for Histograms

Number of Observations	Number of Classes
Less than 50	5-7
50-200	7-9
200-500	9-10
500-1,000	10-11
1,000-5,000	11-13
5,000-50,000	13-17
More than 50,000	17-20

$$\text{Class Interval Width} = \frac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Number of Classes}}$$

Additional Slides From Publisher

We now turn to some additional slides from the textbook publisher on the following topics:

1. Shapes to look for in histograms
2. Time series and cross-section data
3. Plotting time series data

Unfortunately, copies of these slides cannot be provided due to legal restrictions from the publisher. Please take good notes.