Lecture 4

Descriptive Statistics II

Text: Chapter III

STAT 8010 Statistical Methods I
August 28, 2019

Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.1

Whitney Huang
Clemson University

Agenda

Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.2
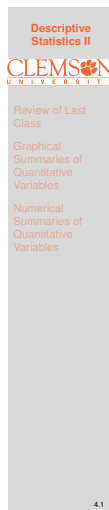
1 **Review of Last Class**

2 **Graphical Summaries of Quantitative Variables**

3 **Numerical Summaries of Quantitative Variables**

Last Lecture

Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.3
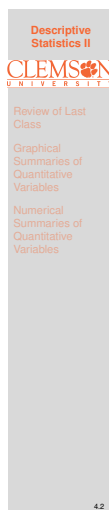
- Summarizing **categorical** variables

  - Frequency Table

  - Bar Chart and Pie Chart

- Summarizing **numerical** variables

  - Mean $\Rightarrow$ A measure of **central tendency**

  - Variance / Standard Deviation $\Rightarrow$ A measure of **spread**
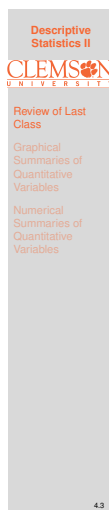
Example: Clemson Fact Sheet 2018

STUDENT DISTRIBUTION						
RACE/ETHNICITY/GENDER			Full Time	Part Time	Total	Age	Average

RACE/ETHNICITY/GENDER			Full Time	Part Time	Total	Age	Average
Non-Resident Alien	6%	Freshman	3,273	10	3,283	16-73	18
Hispanic	4%	Sophomore	4,765	61	4,826	17-79	19
American Indian or Alaskan Native	<1%	Junior	4,786	122	4,908	17-64	20
Asian	2%	Senior	6,121	426	6,547	18-66	22
Black or African American	6%	Uncl. U/G	26	79	105	16-73	24
Native Hawaiian or Pacific Islander	<1%	Master's	1,661	1,629	3,290	19-78	29
White	77%	Doctoral	1,291	292	1,583	20-71	30
Two or More Races (non-Hispanic)	3%	Certificate		94	94	21-56	35
Unknown	1%	Specialists	1	29	30	26-55	36
		Uncl. Grad	16	269	285	23-73	38
51% Male, 49% Female		**Total**	21,940	3,011	24,951	16-79	22

Data source: https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf

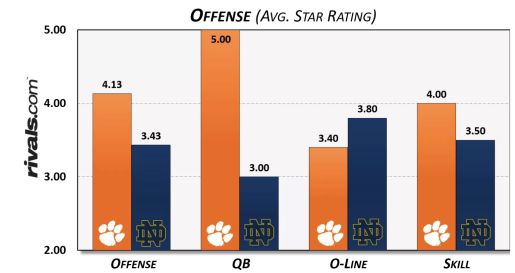Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.4

Notes

Example: Clemson vs. Notre Dame

OFFENSE (AVG. STAR RATING)

rivals.com

	OFFENSE	QB	O-LINE	SKILL
Clemson	4.13	5.00	3.40	4.00
Notre Dame	3.43	3.00	3.80	3.50

Data source: https://www.shakinthesouthland.com/2018/12/24/18132204/cotton-bowl-clemson-vs-notre-dame-preview-depth-chart-statistical-analysis

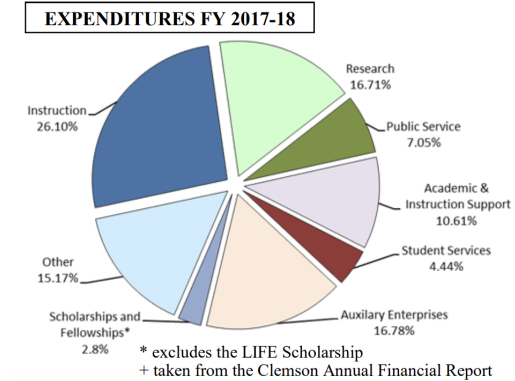Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.5

Notes

Example: Clemson Expenditures 2017-2018

EXPENDITURES FY 2017-18

- Instruction 26.10%
- Research 16.71%
- Public Service 7.05%
- Academic & Instruction Support 10.61%
- Student Services 4.44%
- Auxilary Enterprises 16.78%
- Scholarships and Fellowships* 2.8%
- Other 15.17%

* excludes the LIFE Scholarship
+ taken from the Clemson Annual Financial Report

Data source: https://www.clemson.edu/institutional-effectiveness/documents/oir/minis/F18FactSheetUpdated.pdf

Descriptive Statistics II

CLEMSON
UNIVERSITY

Review of Last Class

Graphical Summaries of Quantitative Variables

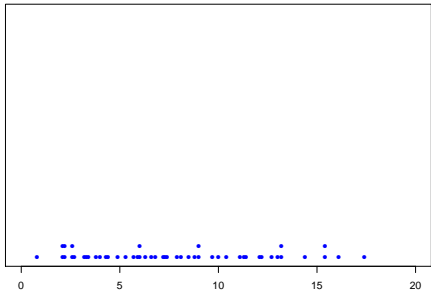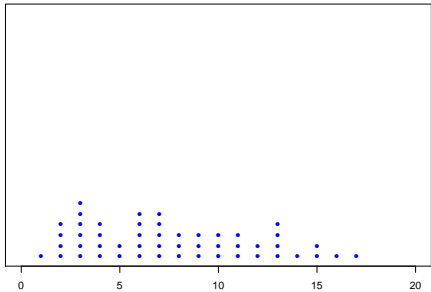Numerical Summaries of Quantitative Variables

4.6

Notes

Example: Murder arrests (per 100,000) in US States in 1973

CLEMS🐾N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

Data: 13.2, 10.0, 8.1, 8.8, 9.0, 7.9, 3.3, 5.9, 15.4, 17.4, 5.3, 2.6, 10.4, 7.2, 2.2, 6.0, 9.7, 15.4, 2.1, 11.3, 4.4, 12.1, 2.7, 16.1, 9.0, 6.0, 4.3, 12.2, 2.1, 7.4, 11.4, 11.1, 13.0, 0.8, 7.3, 6.6, 4.9, 6.3, 3.4, 14.4, 3.8, 13.2, 12.7, 3.2, 2.2, 8.5, 4.0, 5.7, 2.6, 6.8.

Question: How to graphically summarize this data set?

Notes

Dotplot

CLEMS🐾N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

Dotplot of Murder Arrest Rate (per 100,000)

Notes

Dotplot cont'd

CLEMS🐾N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

Rounded Murder Arrest Rate (per 100,000)

Notes

Stem-and-Leaf Plot

```
The decimal point is at the |

 0 | 8
 1 |
 2 | 1122667
 3 | 2348
 4 | 0349
 5 | 379
 6 | 00368
 7 | 2349
 8 | 158
 9 | 007
10 | 04
11 | 134
12 | 127
13 | 022
14 | 4
15 | 44
16 | 1
17 | 4
```

Descriptive
Statistics II

CLEMS☤N
U N I V E R S I T Y

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
Quantitative
Variables

4.10

Notes

Histogram
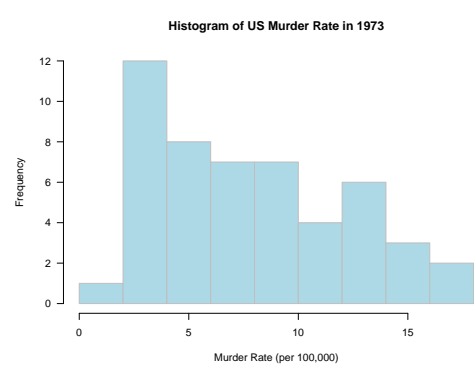
Histogram of US Murder Rate in 1973

Descriptive
Statistics II

CLEMS☤N
U N I V E R S I T Y

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
Quantitative
Variables

4.11

Notes

Histogram
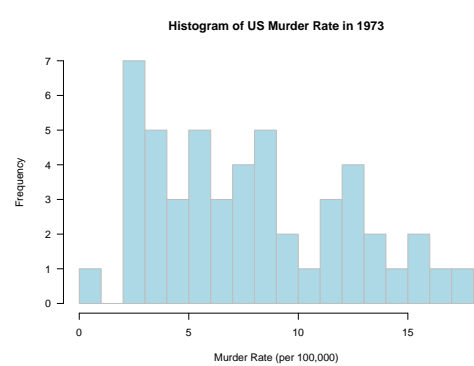
Histogram of US Murder Rate in 1973

Descriptive
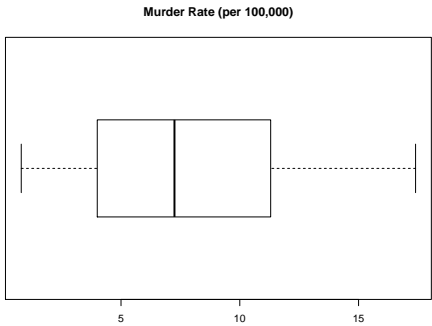Statistics II

CLEMS☤N
U N I V E R S I T Y

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
Quantitative
Variables

4.12

Notes

Box-and-Whisker Plot

Murder Rate (per 100,000)

Descriptive
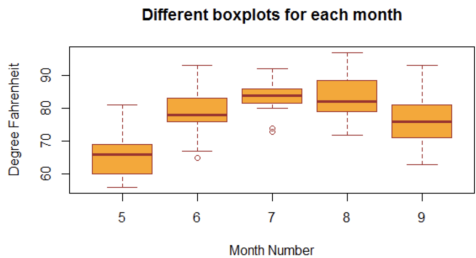Statistics II

CLEMSON
UNIVERSITY

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
Quantitative
Variables

4.13

Notes

Side-by-Side Boxplots

Source: https://www.datamentor.io/r-programming/box-plot/

Descriptive
Statistics II

CLEMSON
UNIVERSITY

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
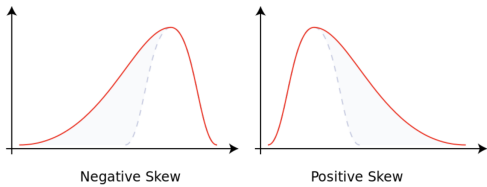Quantitative
Variables

4.14

Notes

Shape of Distributions

Negative Skew   Positive Skew

Source: Skewness - Wikipedia

Descriptive
Statistics II

CLEMSON
UNIVERSITY

Review of Last
Class

Graphical
Summaries of
Quantitative
Variables

Numerical
Summaries of
Quantitative
Variables

4.15

Notes

Shape of Distributions cont'd

t

g/t
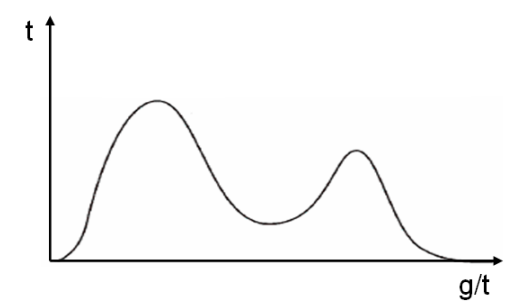
Source: Multimodal distribution - Wikipedia

Descriptive Statistics II

CLEMS😀N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.16

Measures of Center

- A **measure of center** attempts to report a "typical" value for the variable

- When a measure of center is calculated with **sample data** it is a **statistic**

- When a measure of center is calculated with popular (e.g., census data) it is a **parameter**

- **Measures:** Mean, Median, Mode, ...

Descriptive Statistics II

CLEMS😀N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.17

Mean

- The **population mean**, denoted by μ_X, is the sum of all the population values ($\{X_i, \cdots, X_N\}$) divided by the total number (N) of population values. That is,

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$$

- The **sample mean**, denoted by $\bar{X}$ is the sum of all the sample values ($\{X_1, \cdots, X_n\}$) divided by the total number of sample values (n). That is

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Descriptive Statistics II

CLEMS😀N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables

Numerical Summaries of Quantitative Variables

4.18

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13

Descriptive Statistics II

CLEMS⬥N
U N I V E R S I T Y

Review of Last Class

Graphical Summaries of Quantitative Variables
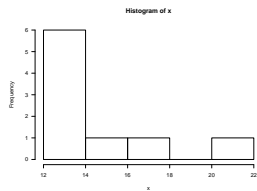
Numerical Summaries of Quantitative Variables

- Plot this "data set" and describe the shape of the distribution

Histogram of x

- Find the mean (both sample and popular means)

$$\bar{X} = \mu_X = \sum_{i=1}^{9} \frac{13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13}{9} = 15$$

Notes

Notes

Notes