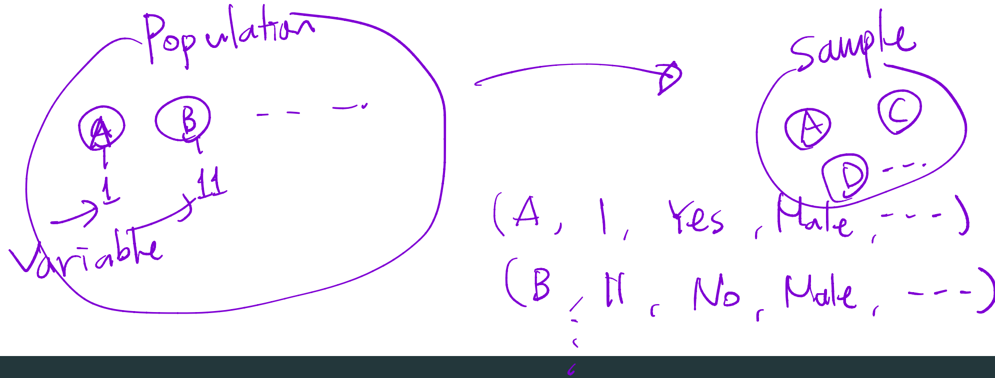
# Chapter 1. Overview and Descriptive Statistics

Math 3670 Spring 2025

Georgia Institute of Technology

# Section 1.
# Populations, Samples, and Processes




## Population and Sample

Population: A well-defined collection of objects.
Ex: All individuals who received a B.S. in engineering during the most recent academic year

Sample: A subset of the population selected in a prescribed way.
Ex: Select a sample of last year's engineering graduates to obtain feedback about the quality of the engineering curricula

Variable: Any characteristic whose value may change from one object to another in the population
Ex: Brand of calculator owned by a student

Univariate, Bivariate, Multivariate Data

## Descriptive Statistics

To summarize and describe important features of the data

1. Graphical methods: Boxplots, Histograms, Scatter plots, etc.
2. Numerical methods: means, standard deviations, correlation coefficients, etc.

Population
All Charities
in US

Sample
60 charities

A: 6.1%
B:
⋮

## Example

Charity is a big business in the United States. The Web site charitynavigator.com gives information on roughly 5500 charitable organizations, and there are many smaller charities that fly below the navigator's radar screen. Some charities operate very efficiently, with fundraising and administrative expenses that are only a small percentage of total expenses, whereas others spend a high percentage of what they take in on such activities. Here is data on fundraising expenses as a percentage of total expenditures for a random sample of 60 charities:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6.1 | 12.6 | 34.7 | 1.6 | 18.8 | 2.2 | 3.0 | 2.2 | 5.6 | 3.8 |
| 2.2 | 3.1 | 1.3 | 1.1 | 14.1 | 4.0 | 21.0 | 6.1 | 1.3 | 20.4 |
| 7.5 | 3.9 | 10.1 | 8.1 | 19.5 | 5.2 | 12.0 | 15.8 | 10.4 | 5.2 |
| 6.4 | 10.8 | 83.1 | 3.6 | 6.2 | 6.3 | 16.3 | 12.7 | 1.3 | 0.8 |
| 8.8 | 5.1 | 3.7 | 26.3 | 6.0 | 48.0 | 8.2 | 11.7 | 7.2 | 3.9 |
| 15.3 | 16.6 | 8.8 | 12.0 | 4.7 | 14.7 | 6.4 | 17.0 | 2.5 | 16.2 |

Stem : Tens digit
leaf : Ones digit



6.1% { Stem: 0
leaf: 6

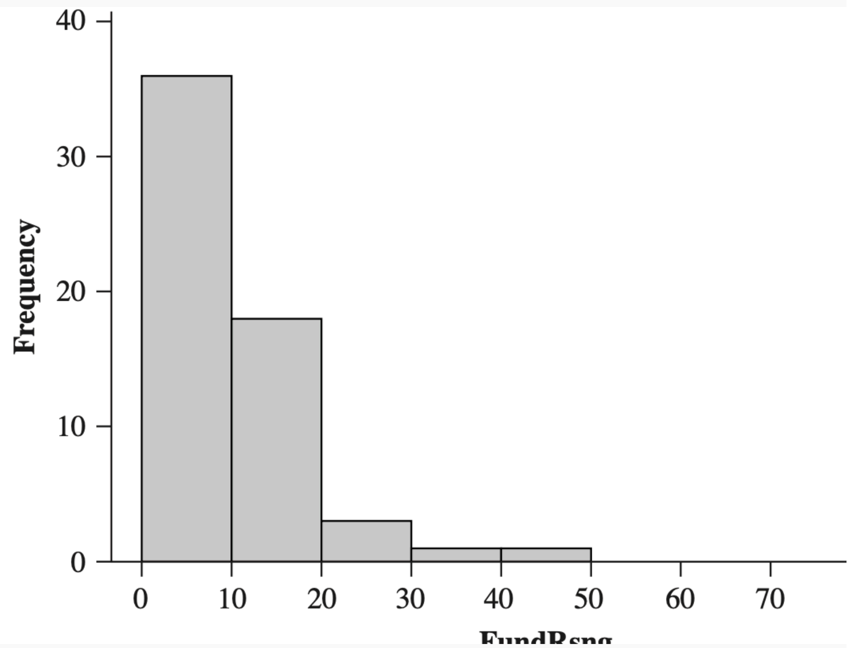
12.6% { Stem : 1
leaf : 2

```
Stem–and–leaf of FundRsng N = 60
Leaf Unit = 1.0
  0 | 0111112222333333344
  0 | 5555666666778888
  1 | 0001222244
  1 | 55666789
  2 | 01
  2 | 6
  3 | 4
  3 |
  4 |
  4 | 8
  5 |
  5 |
  6 |
  6 |
  7 |
  7 |
  8 | 3
```

stem 0

20.x % & 21.x %

26.x %



## Example

Charity is a big business in the United States. The Web site charitynavigator.com gives information on roughly 5500 charitable organizations, and there are many smaller charities that fly below the navigator's radar screen. Some charities operate very efficiently, with fundraising and administrative expenses that are only a small percentage of total expenses, whereas others spend a high percentage of what they take in on such activities. Here is data on fundraising expenses as a percentage of total expenditures for a random sample of 60 charities:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6.1 | 12.6 | 34.7 | 1.6 | 18.8 | 2.2 | 3.0 | 2.2 | 5.6 | 3.8 |
| 2.2 | 3.1 | 1.3 | 1.1 | 14.1 | 4.0 | 21.0 | 6.1 | 1.3 | 20.4 |
| 7.5 | 3.9 | 10.1 | 8.1 | 19.5 | 5.2 | 12.0 | 15.8 | 10.4 | 5.2 |
| 6.4 | 10.8 | 83.1 | 3.6 | 6.2 | 6.3 | 16.3 | 12.7 | 1.3 | 0.8 |
| 8.8 | 5.1 | 3.7 | 26.3 | 6.0 | 48.0 | 8.2 | 11.7 | 7.2 | 3.9 |
| 15.3 | 16.6 | 8.8 | 12.0 | 4.7 | 14.7 | 6.4 | 17.0 | 2.5 | 16.2 |

Mean = Average = $\dfrac{\text{Sum}}{\# \text{ of Data}} = \dfrac{6.1 + 12.6 + \cdots}{60}$

"Typical Number"



Median =

Minimum = 0.8 %

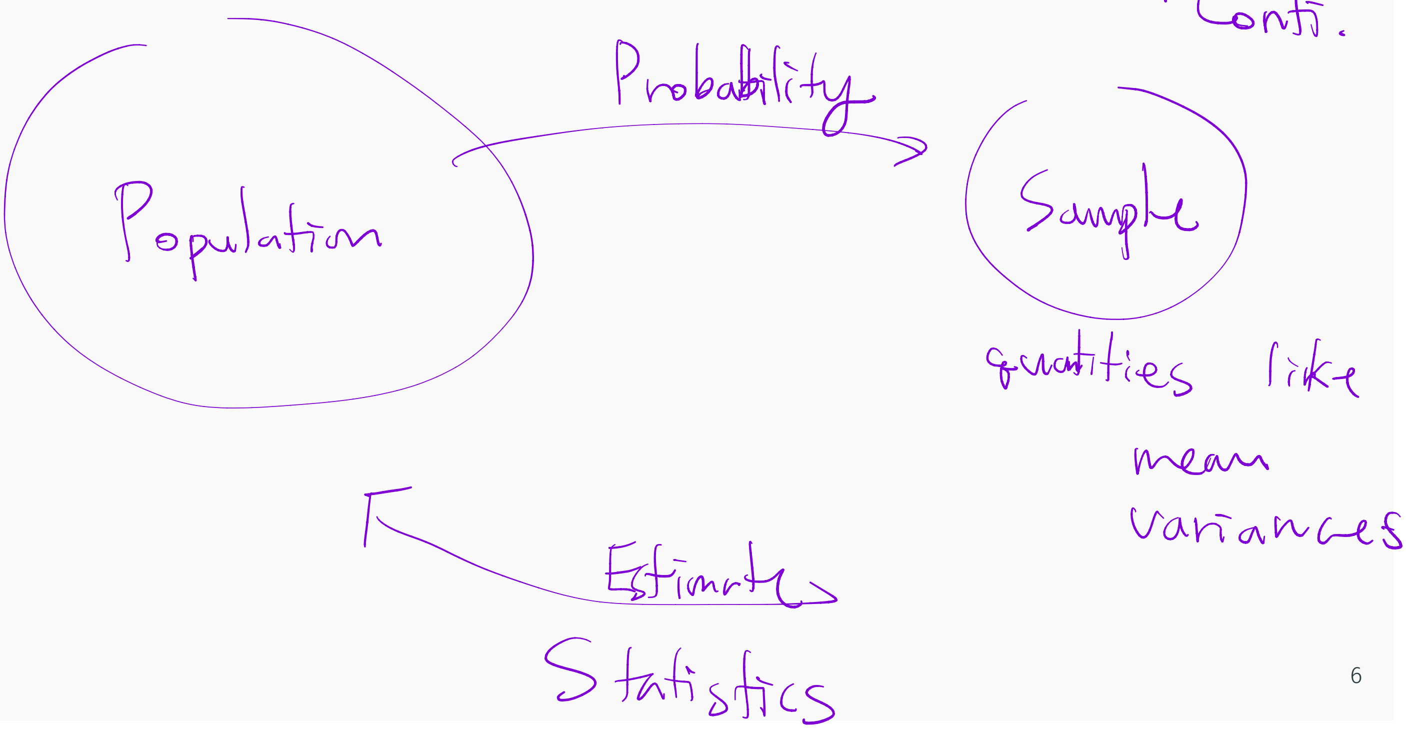
Maximum = 83.1 %

Range = Max − Min

## Overview

Chap 1

Chap 2 – 5 : Prob { Discrete, Conti.



# Section 2.
# Pictorial and Tabular Methods in Descriptive Statistics

## Stem-and-Leaf Displays

Consider a numerical data set $\underbrace{x_1, x_2, \cdots, x_n}$ for which each $x_i$ consists of at least two digits.

Variables

### How to construct Stem-and-Leaf Displays

1. Select one or more leading digits for the stem values. The trailing digits become the leaves.
2. List possible stem values in a vertical column.
3. Record the leaf for each observation beside the corresponding stem value.
4. Indicate the units for stems and leaves someplace in the display.

## Stem-and-Leaf Displays

### Example

4.x %

18.x %

| | |
|---|---|
| 0 | 4 |
| 1 | 1345678889 |
| 2 | 1223456666777889999 |
| 3 | 0112233344555666677777888899999 |
| 4 | 111222223344445566666677788888999 |
| 5 | 00111222233455666667777888899 |
| 6 | 01111244455666778 |

Stem: tens digit
Leaf: ones digit

**Figure 1.4** Stem-and-leaf display for the percentage of binge drinkers at each of the 140 colleges

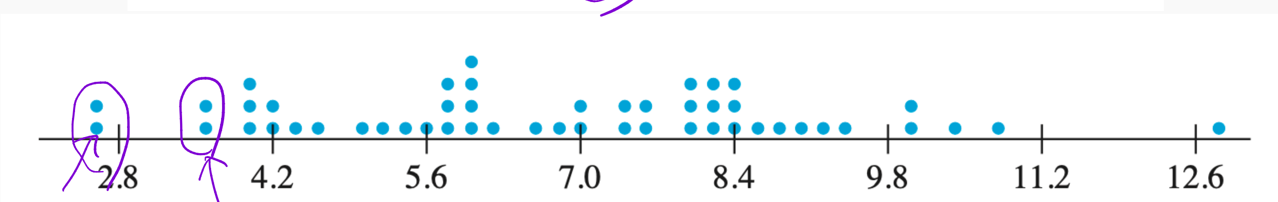
## Stem-and-Leaf Displays

### Information from Stem-and-Leaf Displays

1. identification of a typical or representative value
2. extent of spread about the typical value
3. presence of any gaps in the data
4. extent of symmetry in the distribution of values
5. number and location of peaks
6. presence of any outlying values

## Dotplots

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 10.8 | 6.9 | 8.0 | 8.8 | 7.3 | 3.6 | 4.1 | 6.0 | 4.4 | 8.3 |
| 8.1 | 8.0 | 5.9 | 5.9 | 7.6 | 8.9 | 8.5 | 8.1 | 4.2 | 5.7 |
| 4.0 | 6.7 | 5.8 | 9.9 | 5.6 | 5.8 | 9.3 | 6.2 | 2.5 | 4.5 |
| 12.8 | 3.5 | 10.0 | 9.1 | 5.0 | 8.1 | 5.3 | 3.9 | 4.0 | 8.0 |
| 7.4 | 7.5 | 8.4 | 8.3 | 2.6 | 5.1 | 6.0 | 7.0 | 6.5 | 10.3 |




2.5
2.6

3.5
3.6

# Histograms

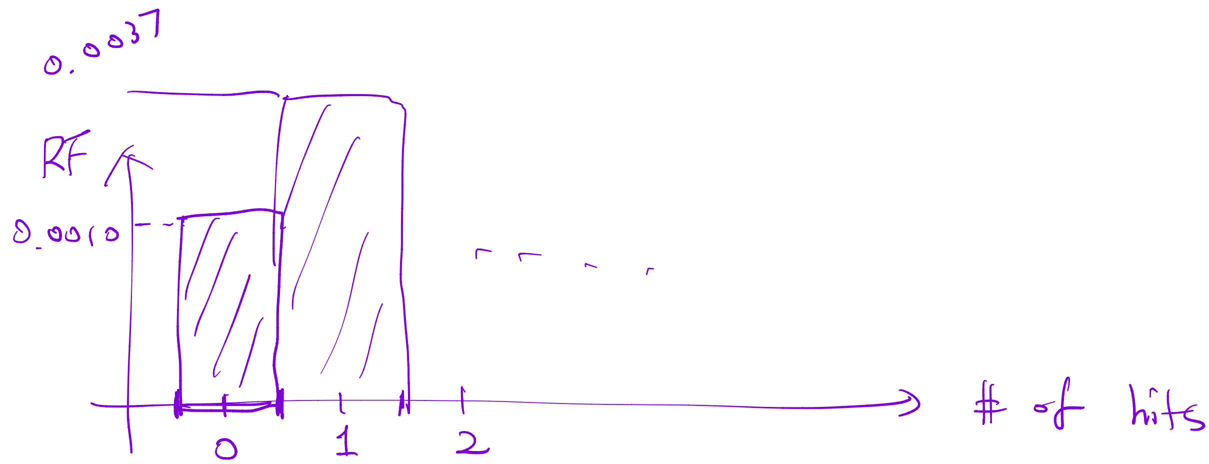
**Definitions**

countable

A numerical variable is discrete if its set of possible values either is finite or else can be listed in an infinite sequence (one in which there is a first number, a second number, and so on).

A numerical variable is continuous if its possible values consist of an entire interval on the number line.

uncountably




## Histograms

### Example (Discrete Data)

variable

| Hits/Game | Number of Games | Relative Frequency | Hits/Game | Number of Games | Relative Frequency |
|---|---|---|---|---|---|
| 0 | 20 | .0010 | 14 | 569 | .0294 |
| 1 | 72 | .0037 | 15 | 393 | .0203 |
| 2 | 209 | .0108 | 16 | 253 | .0131 |
| 3 | 527 | .0272 | 17 | 171 | .0088 |
| 4 | 1048 | .0541 | 18 | 97 | .0050 |
| 5 | 1457 | .0752 | 19 | 53 | .0027 |
| 6 | 1988 | .1026 | 20 | 31 | .0016 |
| 7 | 2256 | .1164 | 21 | 19 | .0010 |
| 8 | 2403 | .1240 | 22 | 13 | .0007 |
| 9 | 2256 | .1164 | 23 | 5 | .0003 |
| 10 | 1967 | .1015 | 24 | 1 | .0001 |
| 11 | 1509 | .0779 | 25 | 0 | .0000 |
| 12 | 1230 | .0635 | 26 | 1 | .0001 |
| 13 | 834 | .0430 | 27 | 1 | .0001 |
| | | | | 19,383 | 1.0005 |

Baseball Games

$$\text{Relative Frequency} = \frac{20}{19\,383}$$

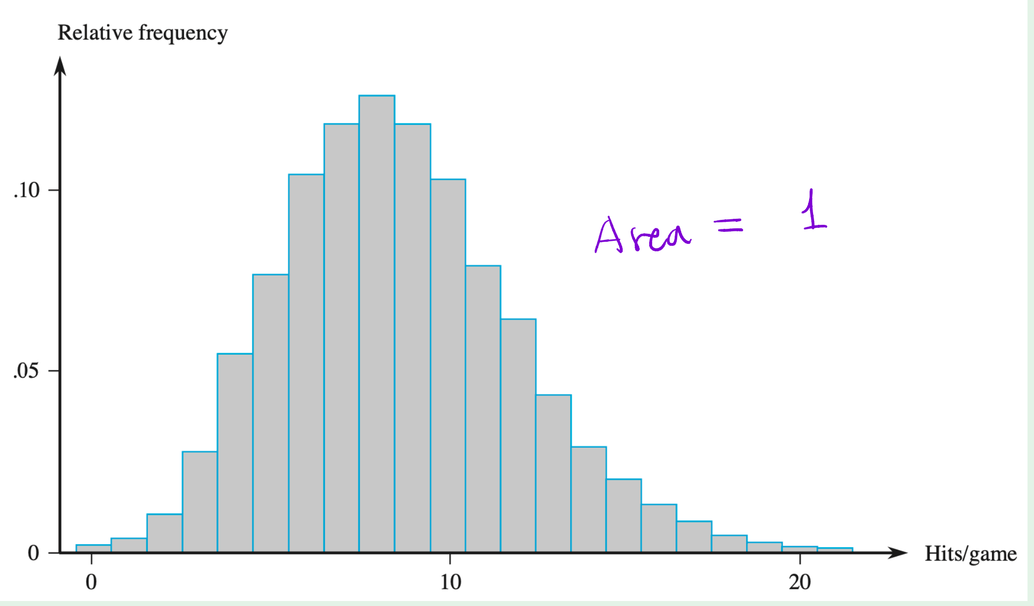
Sum of all RF

$$= 1$$

$$= \frac{\#\text{ of } 0 \text{ hit} + \#\text{ of } 1 \text{ hit} + \cdots}{\#\text{ of Games}}$$

# Histograms

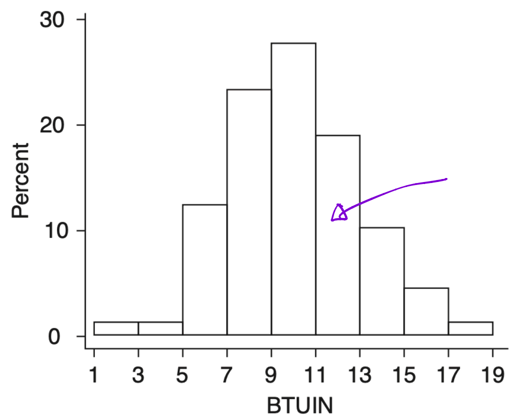
## Example (Discrete Data)



# Histograms

## Example (Continuous Data)

| *Class* | 1−<3 | 3−<5 | 5−<7 | 7−<9 | 9−<11 | 11−<13 | 13−<15 | 15−<17 | 17−<19 |
|---|---|---|---|---|---|---|---|---|---|
| *Frequency* | 1 | 1 | 11 | 21 | 25 | 17 | 9 | 4 | 1 |
| *Relative frequency* | .011 | .011 | .122 | .233 | .278 | .189 | .100 | .044 | .011 |




area = 2

## Example (Continuous Data)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11.5 | 12.1 | 9.9 | 9.3 | 7.8 | 6.2 | 6.6 | 7.0 | 13.4 | 17.1 | 9.3 | 5.6 |
| 5.7 | 5.4 | 5.2 | 5.1 | 4.9 | 10.7 | 15.2 | 8.5 | 4.2 | 4.0 | 3.9 | 3.8 |
| 3.6 | 3.4 | 20.6 | 25.5 | 13.8 | 12.6 | 13.1 | 8.9 | 8.2 | 10.7 | 14.2 | 7.6 |
| 5.2 | 5.5 | 5.1 | 5.0 | 5.2 | 4.8 | 4.1 | 3.8 | 3.7 | 3.6 | 3.6 | 3.6 |

2 2 2 4 8 10

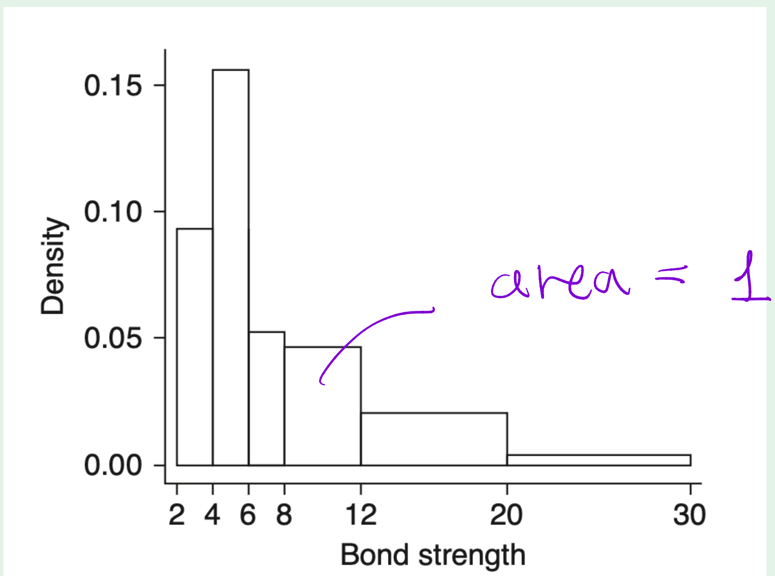
| *Class* | 2–<4 | 4–<6 | 6–<8 | 8–<12 | 12–<20 | 20–<30 |
|---|---|---|---|---|---|---|
| *Frequency* | 9 | 15 | 5 | 9 | 8 | 2 |
| *Relative frequency* | .1875 | .3125 | .1042 | .1875 | .1667 | .0417 |
| *Density* | .094 | .156 | .052 | .047 | .021 | .004 |

$= \frac{0.1875}{2}$ $\qquad$ $= \frac{0.0417}{10}$

$$\text{Density} = \frac{RF}{\text{Size of Interval}}$$

## Histograms

### Example (Continuous Data)




area = 1

Area $= \sum$ density $= \dfrac{RF}{\text{Interval size}}$

Interval size

$= \sum RF$

# Section 3.
# Measures of Locations

# The Mean

**Definition**

Consider a data set $x_1, x_2, \cdots, x_n$.

The sample mean is defined by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

## Example

### Example

People not familiar with classical music might tend to believe that a composer's instructions for playing a particular piece are so specific that the duration would not depend at all on the performer(s).

However, there is typically plenty of room for interpretation, and orchestral conductors and musicians take full advantage of this.

Selected a sample of 12 recordings of Beethoven's Symphony #9 yielding the following durations (min) listed in increasing order:

$$62.3, 62.8, 63.6, 65.2, 65.7, 66.4, 67.4, 68.4, 68.8, 70.8, 75.7, 79.0$$

The sample mean is $\bar{x} = 816.1/12 = 68.01$.

## The Median

$x_1 \quad x_2 \quad x_3 \quad x_4$

### Definition

The sample median is obtained by first ordering the $n$ observations from smallest to largest.

$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$

If the number of the observation is even, $\tilde{x} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$

If the number of the observation is odd, $\tilde{x} = x_{\frac{n+1}{2}}$

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

$x_3 = \tilde{x}$

## Example

### Example

People not familiar with classical music might tend to believe that a composer's instructions for playing a particular piece are so specific that the duration would not depend at all on the performer(s).

However, there is typically plenty of room for interpretation, and orchestral conductors and musicians take full advantage of this.

Selected a sample of 12 recordings of Beethoven's Symphony #9 yielding the following durations (min) listed in increasing order:

$$62.3, 62.8, 63.6, 65.2, 65.7, \boxed{66.4, 67.4}, 68.4, 68.8, 70.8, 75.7, 79.0$$

1 2 3 4 5 6 7 8 9 10 11 12

The sample median $\widetilde{x}$ is

$$\widetilde{x} = \frac{1}{2}(66.4 + 67.4) = 66.9$$

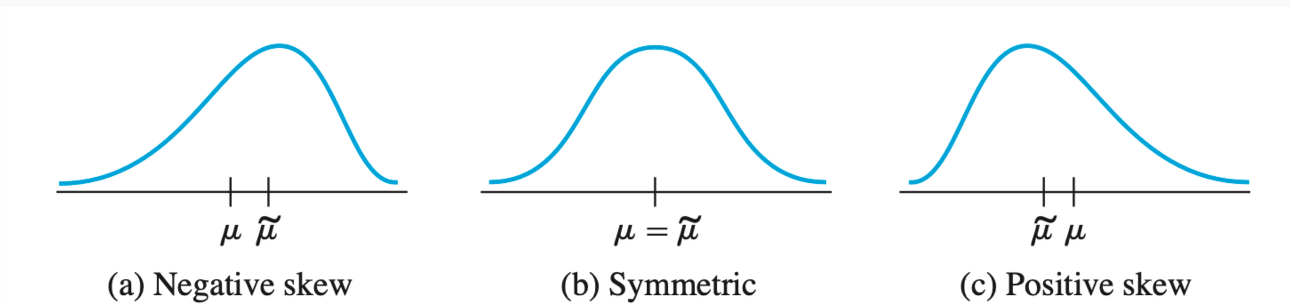
## Population Mean and Median

**Definition**

The population mean is

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

The population median $\widetilde{\mu}$ is the median from the whole population.



$X_1, X_2, \ldots, \ldots, X_{11}$

## Quartiles and Percentiles

**Definition**

Quartiles: divide the number of data points into four equal parts.

Percentiles: divide the number of data points into 100 equal parts.

**Example**

62.3, 62.8, 63.6, 65.2, 65.7, 66.4, 67.4, 68.4, 68.8, 70.8, 75.7, 79.0

1 2 3 4 5 6 7 8 9 10 11 12

median = $q_1$ (1st quartile)

median = $q_3$ (3rd quartile)



$q_1 = \frac{1}{2}(63.6 + 65.2) = 64.4$

$q_2 = \tilde{x} = 66.9$

$q_3 = \frac{1}{2}(68.8 + 70.8)$

## Trimmed Means

**Definition**

A **trimmed mean** a trimming percentage of $\alpha\%$ is the mean of the data set after removing the smallest $\alpha\%$ and the largest $\alpha\%$.

**Example**

Size 12

62.3, 62.8, 63.6, 65.2, 65.7, 66.4, 67.4, 68.4, 68.8, 70.8, 75.7, 79.0

$$\bar{X}_{tr(\alpha)} = \frac{62.8 + \cdots + 75.7}{10}$$

$$\alpha = \frac{1}{12} \times 100 \%$$

## Exercise

The May 1, 2009 issue of The Montclarian reported the fol- lowing home sale amounts for a sample of homes in Alameda, CA that were sold the previous month (1000s of $):

590, 815, 575, 608, 350, 1285, 408, 540, 555, 679.

~~350~~, 408, 540, 555, $\boxed{575, 590}$, 608, 679, 815, ~~1285~~ → 985

The sum is $\boxed{6405}$.

1. Calculate and interpret the sample mean and median.
2. Suppose the 6th observation had been 985 rather than 1285. How would the mean and median change?
3. Calculate a 10% trimmed mean.

① $\bar{x} = \frac{6405}{10} = 640.5 \xrightarrow{-\frac{300}{10}} 610.5$

$\tilde{x} = \frac{1}{2}(575 + 590) = 582.5 \longrightarrow$ no change



② $1285 \longrightarrow 590$

without changing $\tilde{x}$.

③ $\bar{x}_{tr(\alpha)} = \frac{408 + 540 + \cdots + 815}{8} = \frac{6405 - 350 - 1285}{8}$

# Section 4.
# Measures of Variability

## The Sample Variance

**Example**

Consider the two data sets

$$\text{Data 1:} \quad 10, 20, 30, 40, 50, 60, 70,$$
$$\text{Data 2:} \quad 30, 35, 37, 40, 43, 45, 50.$$

$\tilde{x}_1 = 40$, $\bar{x}_1 = 40$

$\tilde{x}_2 = 40$, $\bar{x}_2 = 40$

## The Sample Variance

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

### Definition

$x_1, x_2, \ldots, x_n$

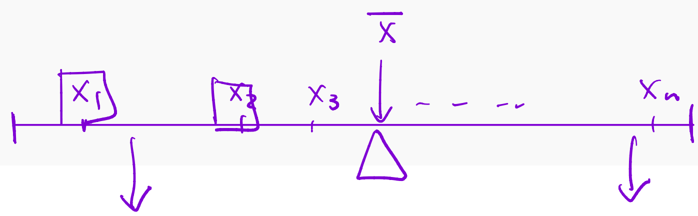
The sample variance is defined by

$$s^2 = \frac{S_{xx}}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

The sample standard deviation is

$$s = \sqrt{s^2}.$$

Note (i) $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$






$$= \underline{\sum_{i=1}^{n} x_i} - \underbrace{\sum_{i=1}^{n} \overline{x}}_{=} = 0$$

$$n \cdot \overline{x} = n \frac{\sum x_i}{n}$$

(ii) Why $(n-1)$?

## The Sample Variance

Suppose the population consists of $x_1, x_2, \cdots, x_N$.

**Definition**

The population variance is defined by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2.$$

# Example

## Example

| Car | $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | |
|---|---|---|---|---|
| 1 | 27.3 | −5.96 | 35.522 | |
| 2 | 27.9 | −5.36 | 28.730 | |
| 3 | 32.9 | −0.36 | 0.130 | |
| 4 | 35.2 | 1.94 | 3.764 | |
| 5 | 44.9 | 11.64 | 135.490 | |
| 6 | 39.9 | 6.64 | 44.090 | |
| 7 | 30.0 | −3.26 | 10.628 | |
| 8 | 29.7 | −3.56 | 12.674 | |
| 9 | 28.5 | −4.76 | 22.658 | |
| 10 | 32.0 | −1.26 | 1.588 | |
| 11 | 37.6 | 4.34 | 18.836 | |
| | $\sum x_i = 365.9$ | $\sum(x_i - \bar{x}) = .04$ | $\sum(x_i - \bar{x})^2 = 314.106$ | $\bar{x} = 33.26$ |

$$s^2 = 31.4106 = \frac{\sum(x_i - \bar{x})^2}{11 - 1}$$

## Computing Formula for $s^2$

$$s^2 = \frac{S_{xx}}{n-1},$$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$n \cdot \bar{x} = \sum x_i$$

**Proposition**

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 =$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} \left( x_i^2 - 2 x_i \cdot \bar{x} + \bar{x}^2 \right)$$

$$= \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} (-2x_i \cdot \bar{x}) + \sum_{i=1}^{n} \bar{x}^2$$

$$= \quad '' \quad - 2 \cdot \bar{x} \cdot \underbrace{\left( \sum_{i=1}^{n} x_i \right)}_{= n \cdot \bar{x}} + n \cdot \bar{x}^2$$



$$= \sum_{i=1}^{n} x_i^2 - 2n \cdot \bar{x}^2 + n \cdot \bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n \cdot \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2$$

## Example

**Example**

Consider the following data:

| | | | | | |
|---|---|---|---|---|---|
| 2.1389 | 2.8132 | 2.4451 | 2.4660 | 2.6038 | 2.4186 |
| 3.8592 | 2.1988 | 2.3529 | 2.2028 | 2.7468 | 1.5104 |
| 2.1987 | 2.5252 | 2.8462 | 2.2722 | 2.2026 | 2.0153 |

Knowing

$$\sum_{i=1}^{18} x_i = 43.8166, \qquad \sum_{i=1}^{18} x_i^2 = 110.5081,$$

find the sample mean and variance.

$$\bar{x} = \frac{\sum_1 x_i}{18} = \frac{43.8166}{18} = *\,*\,*.$$



$$s^2 = \frac{S_{xx}}{18-1}$$

$$S_{xx} = \sum_1 (x_i - \bar{x})^2$$

$$= \sum_1 x_i^2 - \frac{1}{18}\left(\sum x_i\right)^2$$

$$= 110.5081 - \frac{(43.8166)^2}{18}.$$

$$= //$$

## Boxplots

Order the $n$ observations from smallest to largest and separate the smallest half from the largest half.

The median $\tilde{x}$ is included in both halves if $n$ is odd.

Then the lower fourth "$q_1$" is the median of the smallest half and the upper fourth "$q_3$" is the median of the largest half.

A measure of spread that is resistant to outliers is the **fourth spread** $f_s$, given by

$$f_s = q_3 - q_1$$

$x_1, x_2, \ldots, x_{n-1}, x_n$

median $\downarrow$ $q_1$ median. $\downarrow$ $q_3$

$f_s$

## Example

$n = 19$

### Example

Consider the data

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lower → | 40 | 52 | 55 | 60 | 70 | 75 | 85 | 85 | 90 | 90 |
| Upper → | 90 | 92 | 94 | 94 | 95 | 98 | 100 | 115 | 125 | 125 |

Then,

1. The smallest $x_i$: 40
2. The lower fourth: $72.5 = q_1$
3. The median: 90
4. The upper fourth: $96.5 = q_3$
5. The largest $x_i$: 125

$f_s = 96.5 - 72.5$

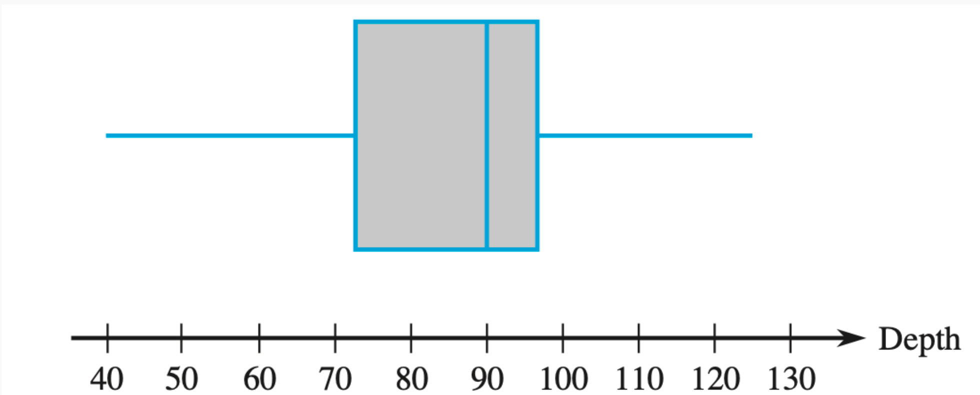
$= 24$

Boxplot



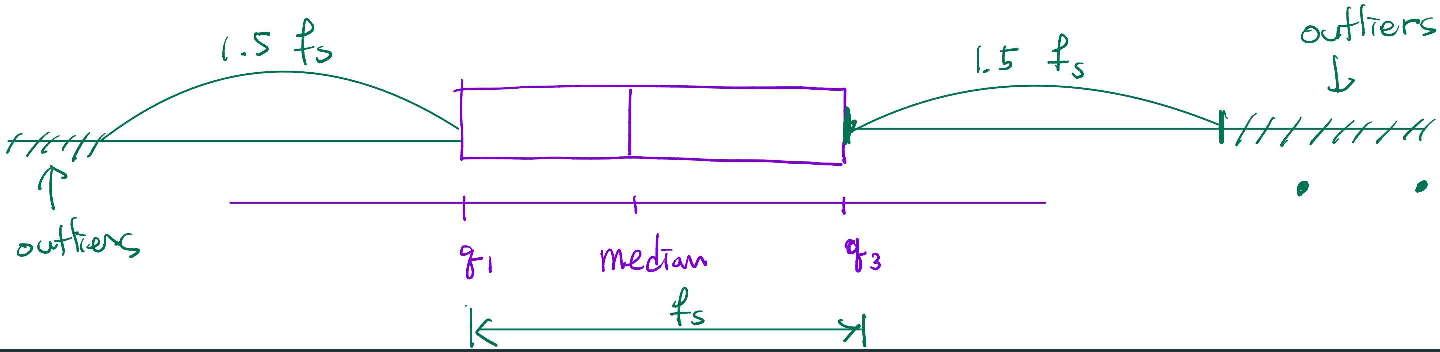

Boxplot.

## Example




| Variable | N | Mean | Median | TrMean | StDev | SE Mean |
|---|---|---|---|---|---|---|
| depth | 19 | 86.32 | 90.00 | 86.76 | 23.32 | 5.35 |

| Variable | Minimum | Maximum | Q1 | Q3 |
|---|---|---|---|---|
| depth | 40.00 | 125.00 | 70.00 | 98.00 |




## Outliers

**Definition**

Any observation farther than $1.5f_s$ from the closest fourth is an outlier.

An outlier is extreme if it is more than $3f_s$ from the nearest fourth, and it is mild otherwise.

## Example

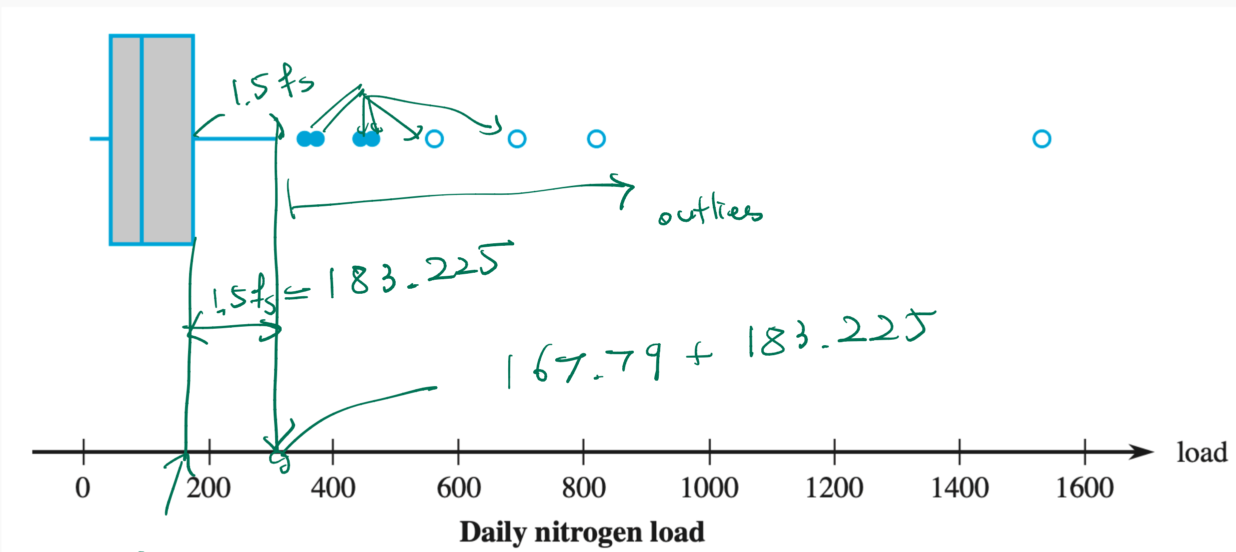
The Clean Water Act and subsequent amendments require that all waters in the United States meet specific pollution reduction goals to ensure that water is "fishable and swimmable." The article "Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads" (*J. of Environ. Engr.,* 2008: 610–618) investigated various techniques for estimating pollutant loads in watersheds; the authors "discuss the imperative need to use sound statistical methods" for this purpose. Among the data considered is the following sample of TN (total nitrogen) loads (kg N/day) from a particular Chesapeake Bay location, displayed here in increasing order.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9.69 | 13.16 | 17.09 | 18.12 | 23.70 | 24.07 | 24.29 | 26.43 |
| 30.75 | 31.54 | 35.07 | 36.99 | 40.32 | 42.51 | 45.64 | 48.22 |
| 49.98 | 50.06 | 55.02 | 57.00 | 58.41 | 61.31 | 64.25 | 65.24 |
| 66.14 | 67.68 | 81.40 | 90.80 | 92.17 | 92.42 | 100.82 | 101.94 |
| 103.61 | 106.28 | 106.80 | 108.69 | 114.61 | 120.86 | 124.54 | 143.27 |
| 143.75 | 149.64 | 167.79 | 182.50 | 192.55 | 193.53 | 271.57 | 292.61 |
| 312.45 | 352.09 | 371.47 | 444.68 | 460.86 | 563.92 | 690.11 | 826.54 |
| 1529.35 | | | | | | | |

## Example

$\tilde{x} = 92.17$ lower $4^{th} = 45.64$ upper $4^{th} = 167.79$

$f_s = 122.15$ $1.5f_s = 183.225$ $3f_s = 366.45$




$q_3 = 167.79$

2:19

## Exercise

The value of Young's modulus (GPa) was determined for cast plates consisting of certain intermetallic substrates, resulting in the following sample observations:

$x_1$ = 116.4 115.9 114.6 115.2 115.8

$\qquad\qquad\ \ \|\ x_2 \quad \|\ x_3 \quad \|\ x_4 \quad \|\ x_5$

1. Calculate $\bar{x}$ and the deviations from the mean.
2. Use the deviations calculated in part (a) to obtain the sample variance and the sample standard deviation.
   $s^2 = \frac{1}{n-1}\left( \sum x_i^2 - \underbrace{\frac{1}{n}\left(\sum x_i\right)^2}_{= n\cdot\bar{x}^2} \right)$
3. Calculate $s^2$ by using the computational formula for the numerator $S_{xx}$.
4. Subtract 100 from each observation to obtain a sample of transformed values. Now calculate the sample variance of these transformed values, and compare it to $s^2$ for the original data.

① $\bar{X} = 115.58$



$x_1 - \bar{x}$, $x_2 - \bar{x}$, ...

② $$s^2 = \frac{S_{xx}}{n-1} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{4} \sum_{i=1}^{5} (x_i - \bar{x})^2$$

$$= 0.482$$

③ $$s^2 = \frac{1}{4} \cdot \Big( \underbrace{\sum x_i^2}_{= 13359.112} - 5 \cdot (115.58)^2 \Big) = 0.482.$$