Probability & Statistics

for Engineers & Scientists

Authors: Walpole, Myers, Myers, YE

Instructor:

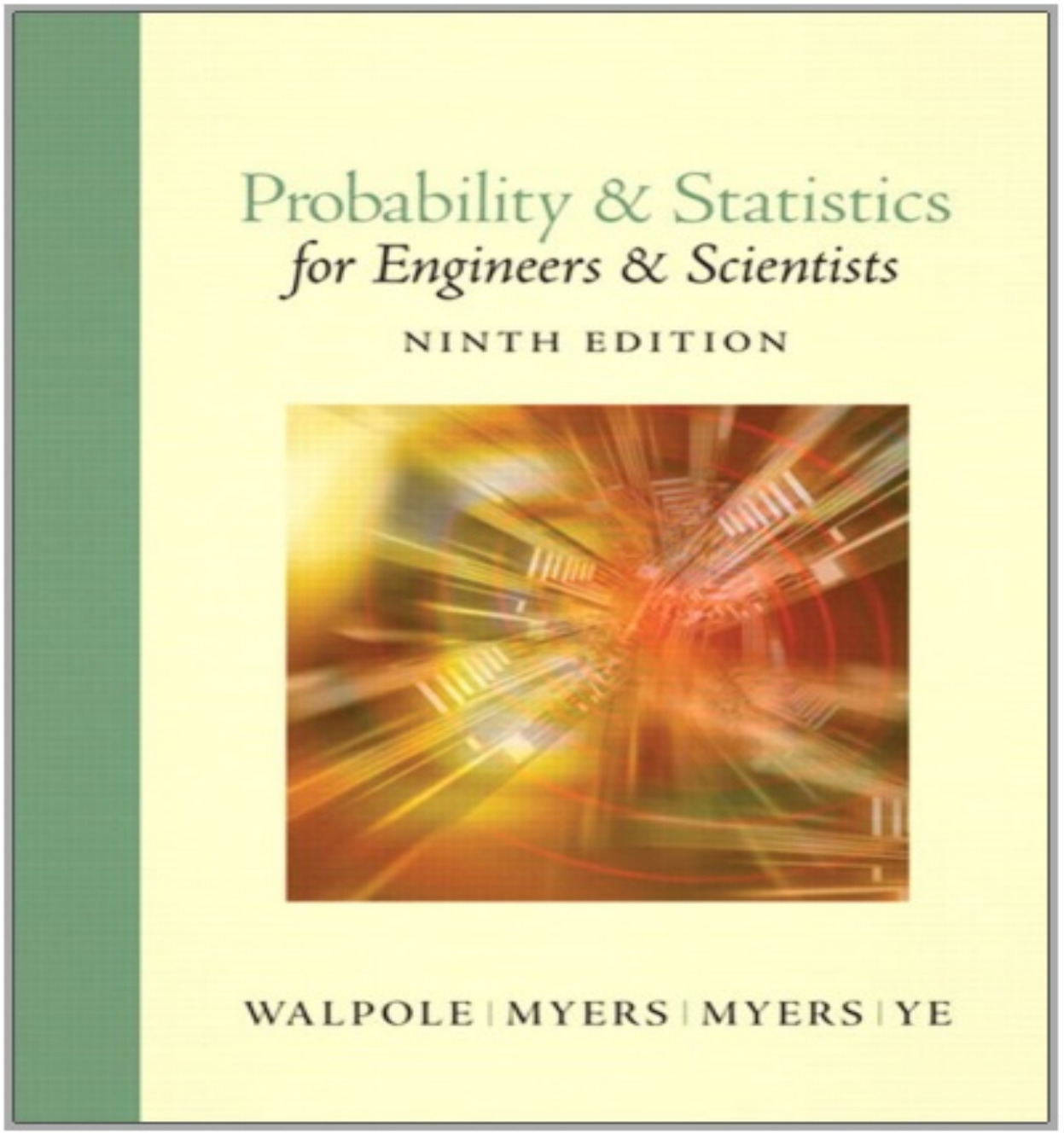
Osama Bin Ajaz

Lecturer, S & H Dept.,

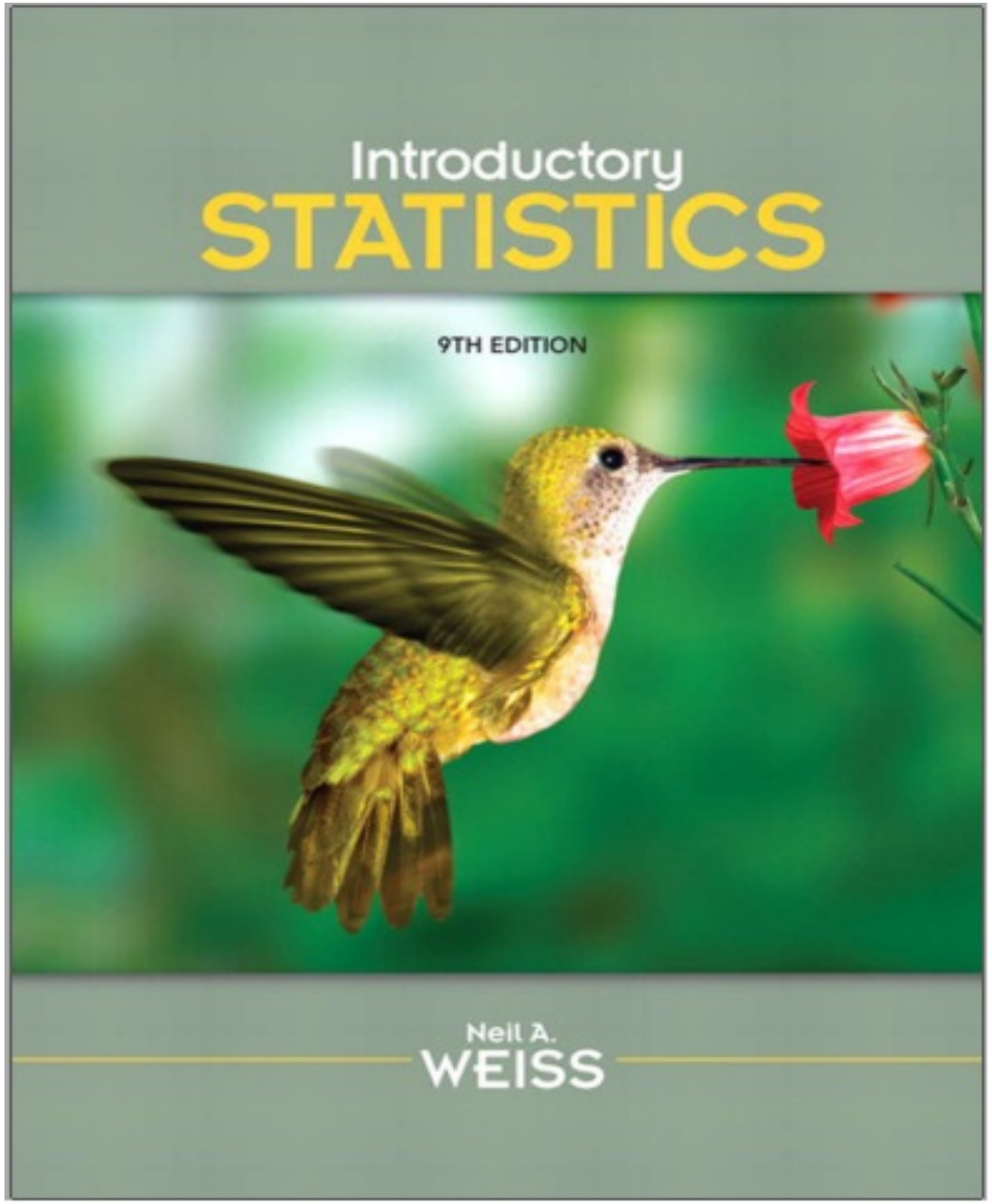
FAST-NU, Main Campus, Karachi

osama.ajaz@nu.edu.pk

Text Book


Probability & Statistics
for Engineers & Scientists
NINTH EDITION
WALPOLE | MYERS | MYERS | YE

Reference Book



Course Outline:

MT2005 – Probability & Statistics
(3 Credit Hours)

Week	Topics
01	**Intro. To Statistics**, Measures of Central Tendency & Dispersion
02	Bar Chart, Histogram, Stem-Leaf Plot, Box Plot, Dot Plot, Frequency Curves, Ogive, Skewness & Kurtosis
03	**Introduction to Probability**: Sample Space, Tree Diagram, Event, Set Theory, Venn Diagram
04	Counting techniques, Kinds of Events, Additive rules
05	Conditional Probability, Independence, Multiplicative rules, Bayes' Theorem.
06	**1st Mid-Term Examination**

Course Outline (Contd.)

MT206 – Probability & Statistics
(4 Credit Hours)

Week	Topics
07	**Random Variables & Probability Distributions:** PMF, PDF, CDF, Joint & Marginal Probability Distributions, Mathematical Expectation
08	**Discrete Distributions:** Binomial & Multinomial, Poisson, Geometric, Hypergeometric, and Discrete uniform.
09	**Continuous Distributions:** Normal, Exponential, Uniform, Chi-Square
10	**Testing of Hypothesis:** z-test, t-test
11	Goodness of Fit Test, Chi-Square test of Independence
12	**2nd Mid-Term Examination**

Course Outline (Contd.)

MT206 – Probability & Statistics
(4 Credit Hours)

Week	Topics
13	Correlation & Regression
14	Polynomial regression
15	ANOVA
16	**Final Examination**

- ***Note:*** *The above course outline & schedule is tentative.*

Marks Distribution

S. No.	Particulars	% Marks
01	Assignments	10
02	Quizzes *(Unannounced)*	10
03	1st Mid Term	15
04	2nd Mid Term	15
05	Final Exam	50
Total		100

Important Instructions

- Be in the classroom on time.
- All students are required to maintain 80% of attendance. In case students fail to maintain 80% of attendance, they become ineligible to take the final exam.
- **Turn off your cell phones or any other electronic devices before entering the class.**
- Maintain the decorum of the classroom all the time.
- Avoid conversation with your classmates while the lecture is in progress.
- Submit your assignments on time otherwise marks will be deducted after deadline.

Important Instructions (Contd.)

- Assignment should include a title page consisting of your complete Name, Roll No, Subject **Name** and date etc.
- Assignment should be submitted in the Holes clip punch **folder** (snap attached).
- Incomplete assignments lead to reduction in marks.
- Avoid plagiarism.
- For Quizzes bring your own **loose pages**.
- ***Violation of any instructions leads to a reduction in marks.***

Please do not ask for lectures slides.

Do follow recommended books.

Thank you

Introduction to Statistics

Introduction to Statistics

- What does the word *statistics* bring to mind?
- If you read newspapers, surf the Web, watch the news on television, or follow sports, you see and hear the following statements:

- 10 year inflation rates.
- Population growth rate, Birth and Death rates etc.
- Exchange rates, Stock prices etc.
- Census & Survey
- Average Run rate of a player
- Ranking of Cricket players.
- Average salary of programmers.

Definition of Statistics

The Science of collection, presentation, analysis and Interpretation of numerical (or non-numerical) facts & data.

Descriptive & Inferential Statistics

- **Descriptive statistics** consists of methods for organizing and summarizing information.
  - *It includes the construction of graphs, charts, tables, and computation of averages, variations, and percentiles.*

- **Inferential Statistics** consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample.
  - *When an inference is made about the population, then the study becomes inferential.*

Further Classifying Statistical Studies

- **Observational Study:** researchers simply observe characteristics and take measurements, as in a sample survey.

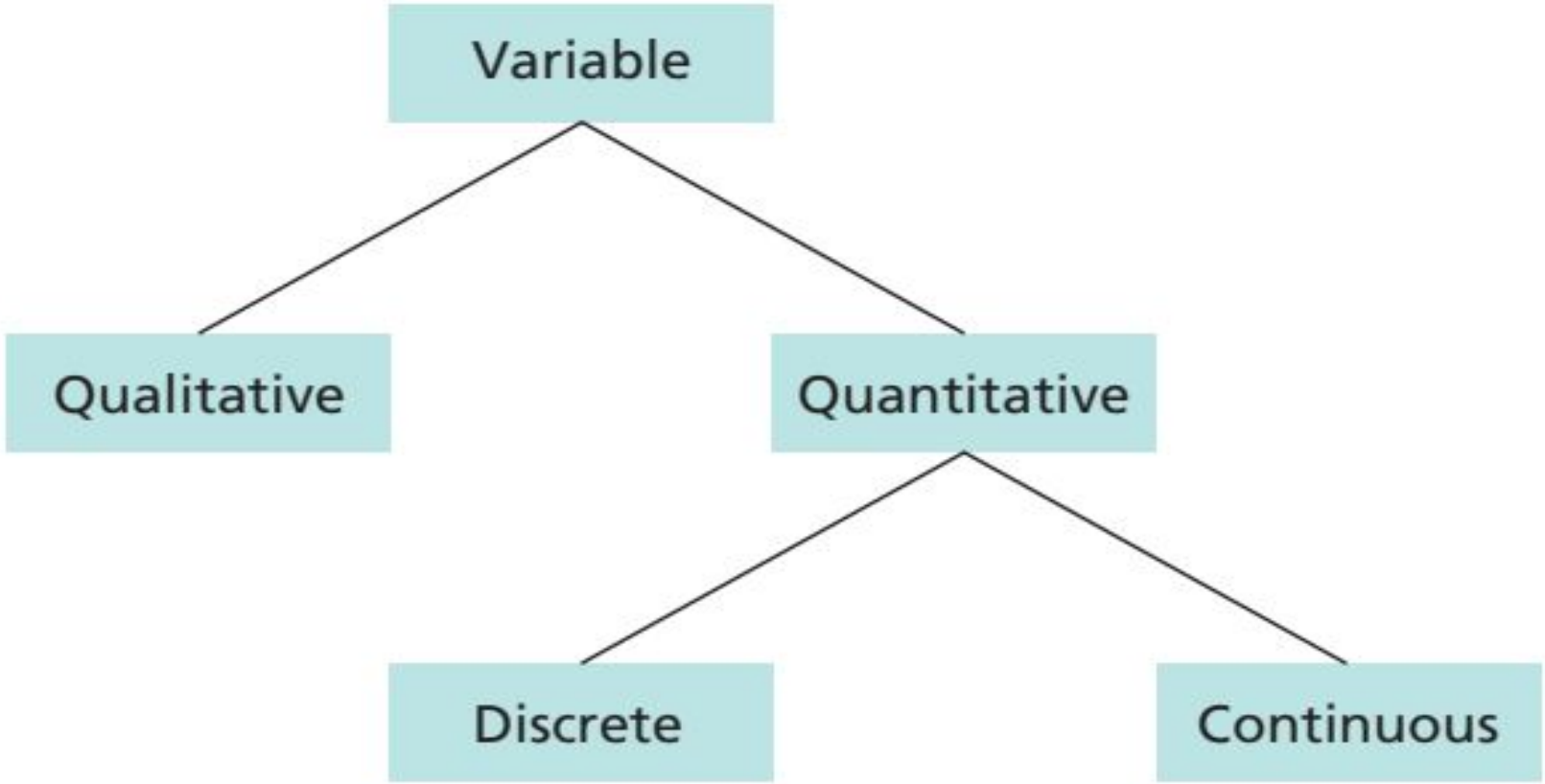
- **Designed experiment:** researchers impose treatments and controls and then observe characteristics and take measurements.

Observational studies can reveal only association, whereas designed experiments can help establish causation.

Statistical Terminologies

- **Population:** The collection of all individuals or items under consideration in a statistical study.
- **Sample:** That part of the population from which information is obtained.
- **Random Sample:** it is a sample selected in such a way that every member of the population has an equal chance of being selected.
- **Parameter:** it describes characteristics of a population.
- **Statistic:** it describe characteristics of a sample.

Statistical Terminologies



Types of Data

- **Primary Data:** it is also called Raw Data or First hand information such as Census & Survey questionnaires etc.

- **Secondary Data**: that is already collected by someone else such as published reports of research organizations.

Measurement Scales

- **Nominal Scale:** **Classification** into mutually exclusive qualitative categories such as Male & Female, Muslims & Non-Muslims, Profession, Blood group etc.

- **Ordinal Scale:** It includes the characteristic of a nominal scale and in addition has the property of **ordering or ranking** of measurements. For example: Performance of students, Customer ratings (strongly agree, agree, neutral, disagree, strongly disagree) etc.

- **Interval Scale:** it includes those quantitative variables for which zero is not meaningful. For example: Temperature, IQ score etc.

- **Ratio Scale:** It is the strongest scale and possess **absolute zero**. For example: Area, Volume, length, distance, weight, money etc.

Classify following variables as discrete or continuous

1. The Life time of a laptop component. **Continuous**
2. The number of heads in toss of 4 coins. **discrete**
3. The amount of milk produced by a cow. **Continuous**
4. The number of enrollment in statistics class. **discrete**
5. The age of a programmer. **Continuous**
6. Processing speed of a software. **Continuous**
7. The number software crashes in 3 hours. **discrete**

Classify as Qualitative or Quantitative

1. Average Salary of Software Engineers in East Asia. **Quantitative**
2. Marks of 40 students of a statistics class. **Quantitative**
3. Program preferences of a talk show. **Qualitative**
4. Religion of peoples in a country. **Qualitative**

Classify scales of measurements

(i) Rankings of golfers in a tournament
(ii) Temperatures inside 10 pizza ovens.
(iii) Weights of selected cell phones
(iv) Salaries of the coaches in the NFL.
(v) Ratings of textbooks (poor, fair, good, excellent).
(vi) Age of children in a day care center
(vii) Categories of magazines in a physician's office (sports, women's, health, men's, news)

Frequency Distribution

- A **frequency distribution of qualitative data** is a listing of the distinct values and their frequencies.

To Construct a Frequency Distribution of Qualitative Data

Step 1 **List the distinct values of the observations in the data set in the first column of a table.**

Step 2 **For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.**

Step 3 **Count the tallies for each distinct value and record the totals in the third column of the table.**

Example # 01: Qualitative Freq. distribution

- The following data represents the political party affiliation of the students in introductory statistics course of a particular college in the USA. Organize these data into a frequency distribution.

D	R	O	R	R	R	R	R
D	O	R	D	O	O	R	D
D	R	O	D	R	R	O	R
D	O	D	D	D	R	O	D
O	R	D	R	R	R	R	D

Example # 02: Bar Chart for Qualitative Freq. distribution

Example # 03: Qualitative Freq. distribution (Contd.)

- The following table provides data on college for the students in one section of the course Introduction to Computer Science during one semester at Arizona State University. In the table, we use the abbreviations BUS for Business, ENG for Engineering and Applied Sciences, and LIB for Liberal Arts and Sciences.
- Construct Frequency distribution and simple Bar chart.

ENG	ENG	BUS	BUS	ENG
LIB	LIB	ENG	ENG	ENG
BUS	BUS	ENG	BUS	ENG
LIB	BUS	BUS	BUS	ENG
ENG	ENG	LIB	ENG	BUS

Example # 04: Quantitative Freq. distribution

- Use single-value grouping to organize the following data into frequency and distribution.

Number of TV sets in each of 50 randomly selected households

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

Example # 05: Group Frequency Distribution

- The following data displays the number of days to maturity for 40 short-term investments. The data are from *BARRON'S* magazine. Use limit grouping, with grouping by 10s, to organize these data into frequency distribution.

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Example # 06: Histogram

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

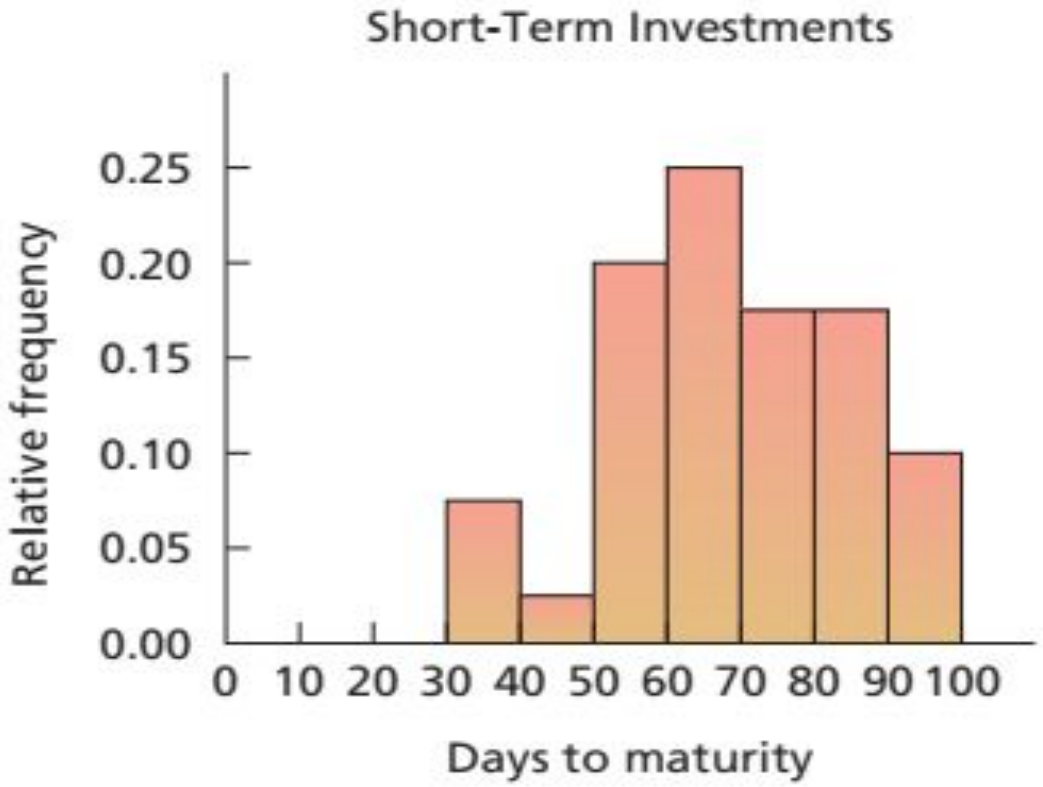
Example # 07: Stem & Leaf Plot

- Reconsider example # 05 and construct Stem & Leaf Plot.

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Example # 07: Stem & Leaf Plot (Contd.)

- Stem & Leaf Plot is similar to Histogram:



Stems	Leaves
3	6 8 9
4	7
5	0 1 1 3 5 5 6 7
6	0 2 3 4 4 5 6 7 8 9
7	0 0 0 1 5 8 9
8	0 1 3 5 6 7 9
9	5 8 9 9

Example # 08: Dotplot

- Dot Plots are useful for showing the relative positions of the data in a data set or for comparing two or more data sets.
- **Construct a dotplot for the following exam scores of the students in an introductory statistics class.**

88	82	89	70	85
63	100	86	67	39
90	96	76	34	81
64	75	84	89	96

2.66 Stressed-Out Bus Drivers. Frustrated passengers, congested streets, time schedules, and air and noise pollution are just some of the physical and social pressures that lead many urban bus drivers to retire prematurely with disabilities such as coronary heart disease and stomach disorders. An intervention program designed by the Stockholm Transit District was implemented to improve the work conditions of the city's bus drivers. Improvements were evaluated by G. Evans et al., who collected physiological and psychological data for bus drivers who drove on the improved routes (intervention) and for drivers who were assigned the normal routes (control). Their findings were published in the article "Hassles on the Job: A Study of a Job Intervention With Urban Bus Drivers" (*Journal of Organizational Behavior*, Vol. 20, pp. 199–208). Following are data, based on the results of the study, for the heart rates, in beats per minute, of the intervention and control drivers.

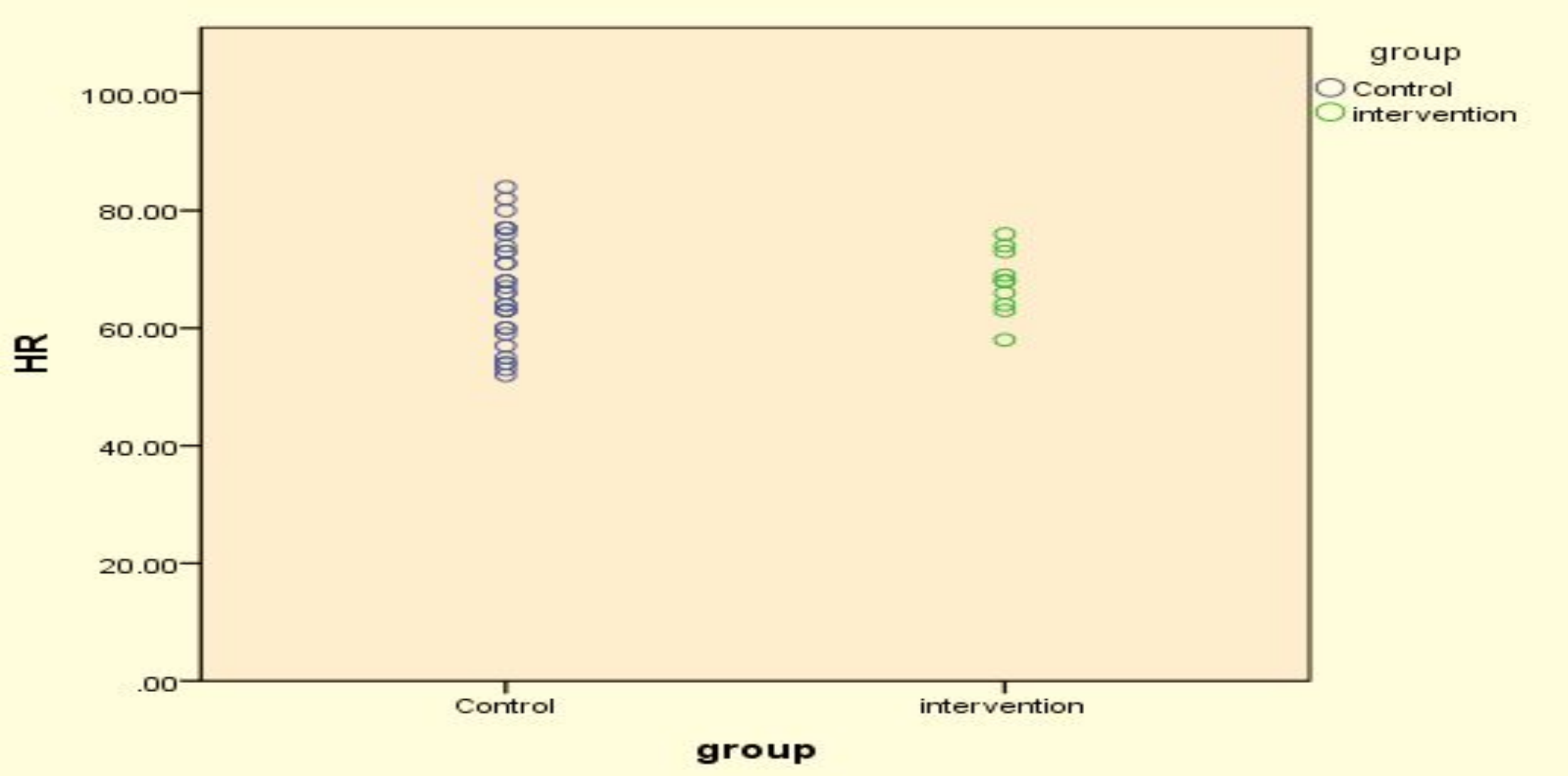
Stressed-Out Bus Drivers (Contd.)

Intervention		Control						
68	66	74	52	67	63	77	57	80
74	58	77	53	76	54	73	54	
69	63	60	77	63	60	68	64	
68	73	66	71	66	55	71	84	
64	76	63	73	59	68	64	82	

a. Obtain dotplots for each of the two data sets, using the same scales.

b. Use your result from part (a) to compare the two data sets.

Stressed-Out Bus Drivers (Contd.)



Convert the following into Freq. Distribution

2.107 Stays in Europe and the Mediterranean. The Bureau of Economic Analysis gathers information on the length of stay in Europe and the Mediterranean by U.S. travelers. Data are published in *Survey of Current Business*. The following stem-and-leaf diagram portrays the length of stay, in days, of a sample of 36 U.S. residents who traveled to Europe and the Mediterranean last year.

0	1 1 1 2 3 3 3 5 5 6 8
1	0 0 0 1 2 2 2 3 4 5 6 7 8
2	0 1 1 1 7
3	1 2
4	1 4 8
5	6
6	4

Example # 09: Frequency Polygon & OGIVE

- Reconsider example # 06 and construct frequency polygon & OGIVE:

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

Measures of Central Tendency
(Mean, Median, Mode)

Example # 10: The Mean

- Reconsider Example # 09:

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

Example # 11: The Median

- Reconsider Example # 10 and calculate Median.

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

Example # 12: The Mode

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

49	57	38	73	81
74	59	76	65	69
54	56	69	68	78
65	85	49	69	61
48	81	68	37	43
78	82	43	64	67
52	56	81	77	79
85	40	85	59	80
60	71	57	61	69
61	83	90	87	74

Solution

L	U	f	CF	x	f.x	f.(x-66.44)^2
36	42	3	3	39	117	2258.8608
43	49	5	8	46	230	2088.968
50	56	4	12	53	212	722.5344
57	63	8	20	60	480	331.7888
64	70	10	30	67	670	3.136
71	77	6	36	74	444	342.9216
78	84	9	45	81	729	1907.9424
85	91	5	50	88	440	2324.168
	Sum	**50**			**3322**	**9980.32**

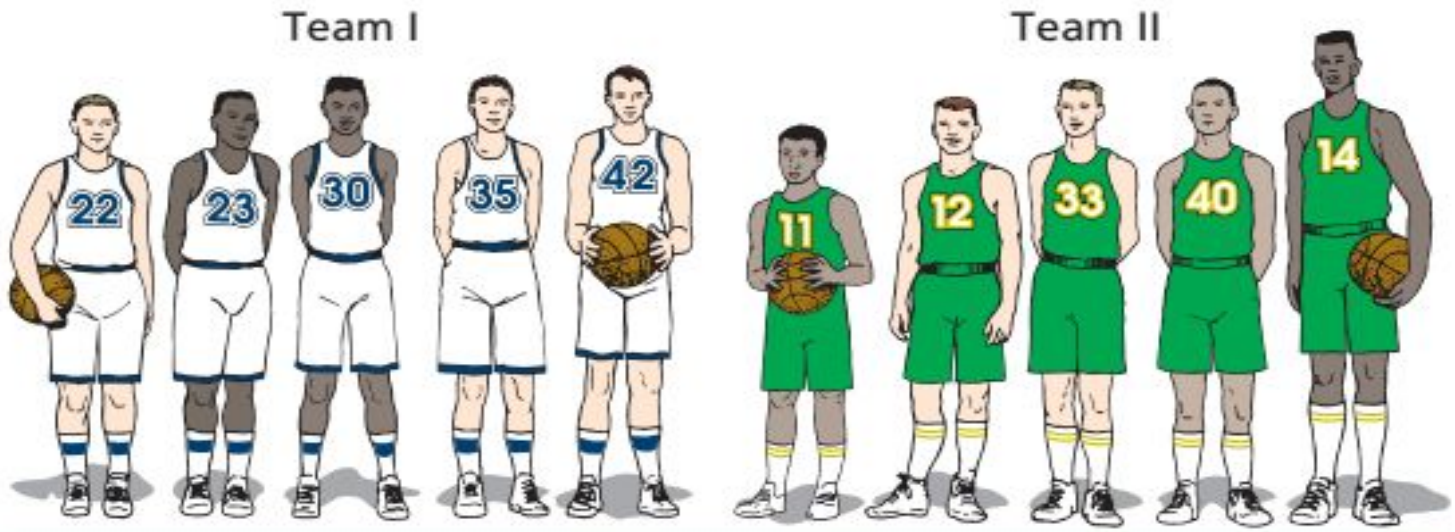
	mean	66.44
	Variance	199.6064
	SD	14.1282129
Q2	**median**	67
fm	**Mode**	65.8333333
sum(f)/4=12.5	**Q1**	56.9375
37.5	**Q3**	78.6666667

Results from ungroup data		
x		
N	Valid	50
	Missing	0
Mean		66.2600
Median		68.0000
Mode		69.00
Variance		196.237
Percentiles	25	56.7500
	50	68.0000
	75	78.2500

Measures of Variation

- Two or more data sets can have same mean, median or mode, but those datasets may differ in other aspects.

Five starting players on two basketball teams

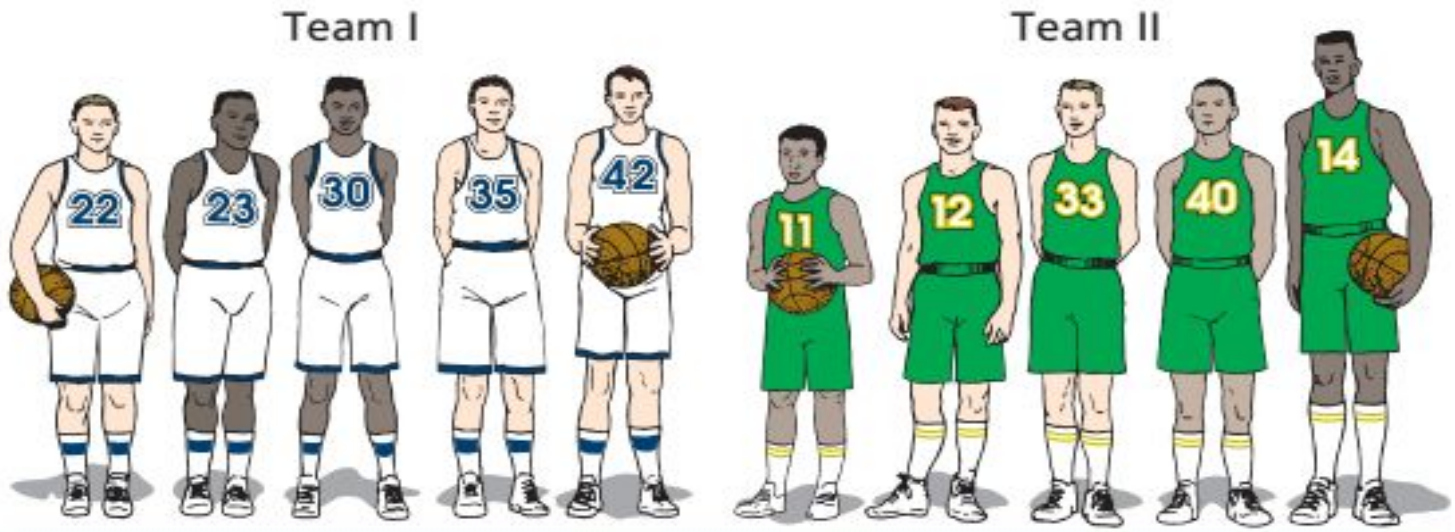


Feet and inches	6'	6'1"	6'4"	6'4"	6'6"	5'7"	6'	6'4"	6'4"	7'
Inches	72	73	76	76	78	67	72	76	76	84

Measures of Variation

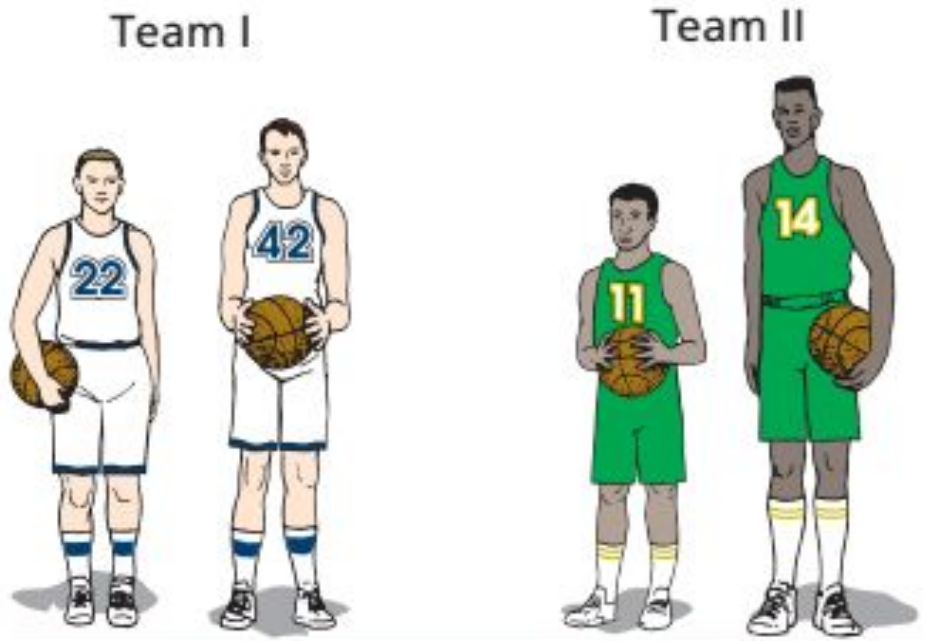
- Two or more data sets can have same mean, median or mode, but those datasets may differ in other aspects.

Five starting players on two basketball teams



	Team I					Team II				
Feet and inches	6'	6'1"	6'4"	6'4"	6'6"	5'7"	6'	6'4"	6'4"	7'
Inches	72	73	76	76	78	67	72	76	76	84

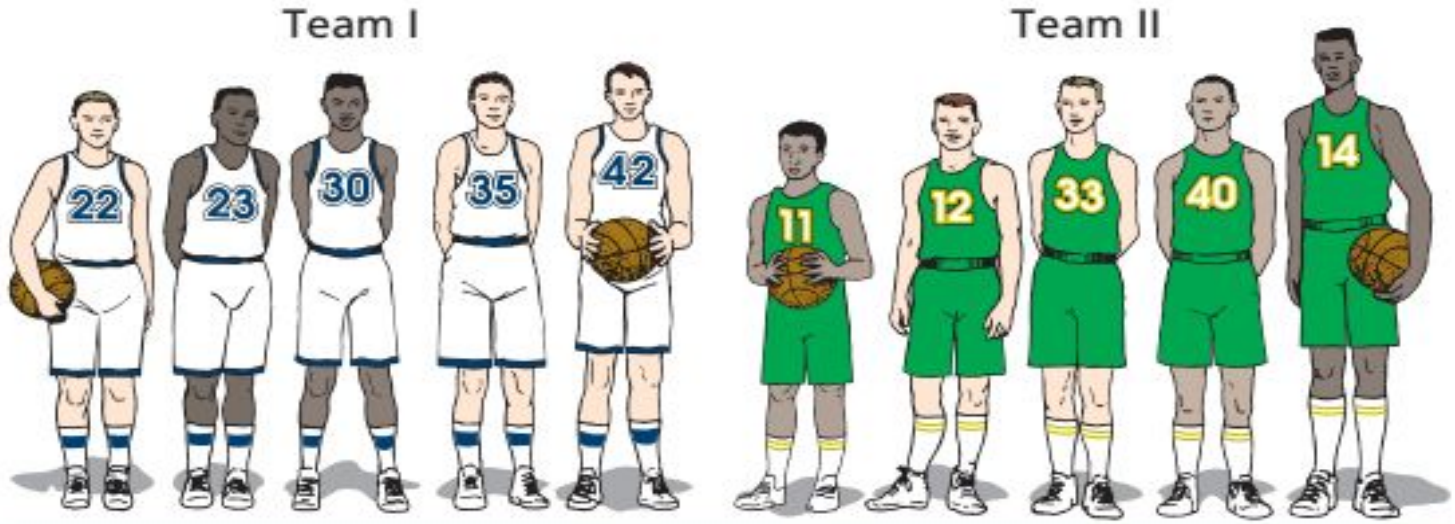
Measures of Variation: RANGE



	Team I		Team II	
Feet and inches	6'	6'6"	5'7"	7'
Inches	72	78	67	84

Measures of Variation: The Standard Deviation

Five starting players on two basketball teams



	Team I					Team II				
Feet and inches	6'	6'1"	6'4"	6'4"	6'6"	5'7"	6'	6'4"	6'4"	7'
Inches	72	73	76	76	78	67	72	76	76	84

Measures of Variation: The Coefficient of Variation

- A statistic that allows you to compare standard deviations when the units are different is called the ***coefficient of variation***.

Example # 16

Sales of Automobiles

The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is $5225, and the standard deviation is $773. Compare the variations of the two.