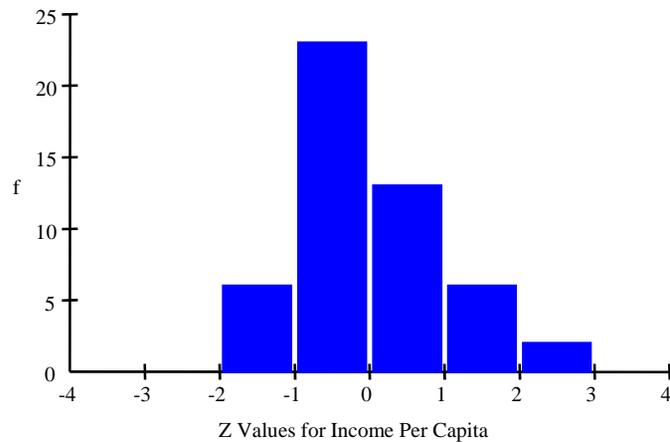


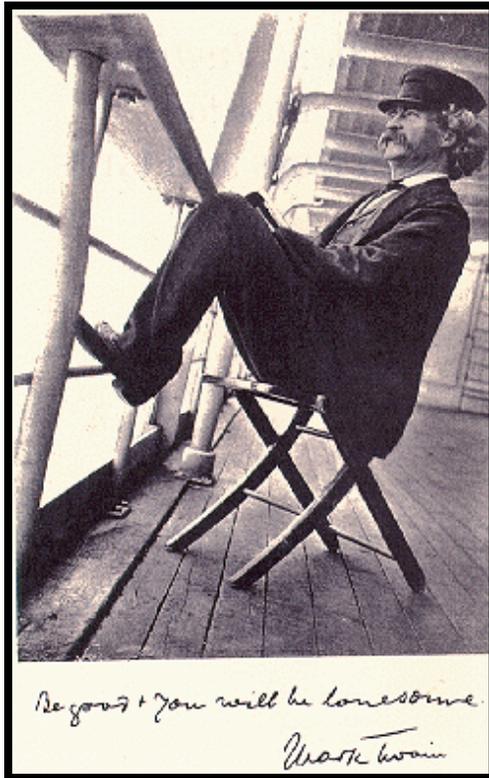
Visualizing and Quantifying Data

MA 206



μ_1

Introduction



Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: **"There are three kinds of lies: lies, damned lies and statistics."**

- *Autobiography of Mark Twain*

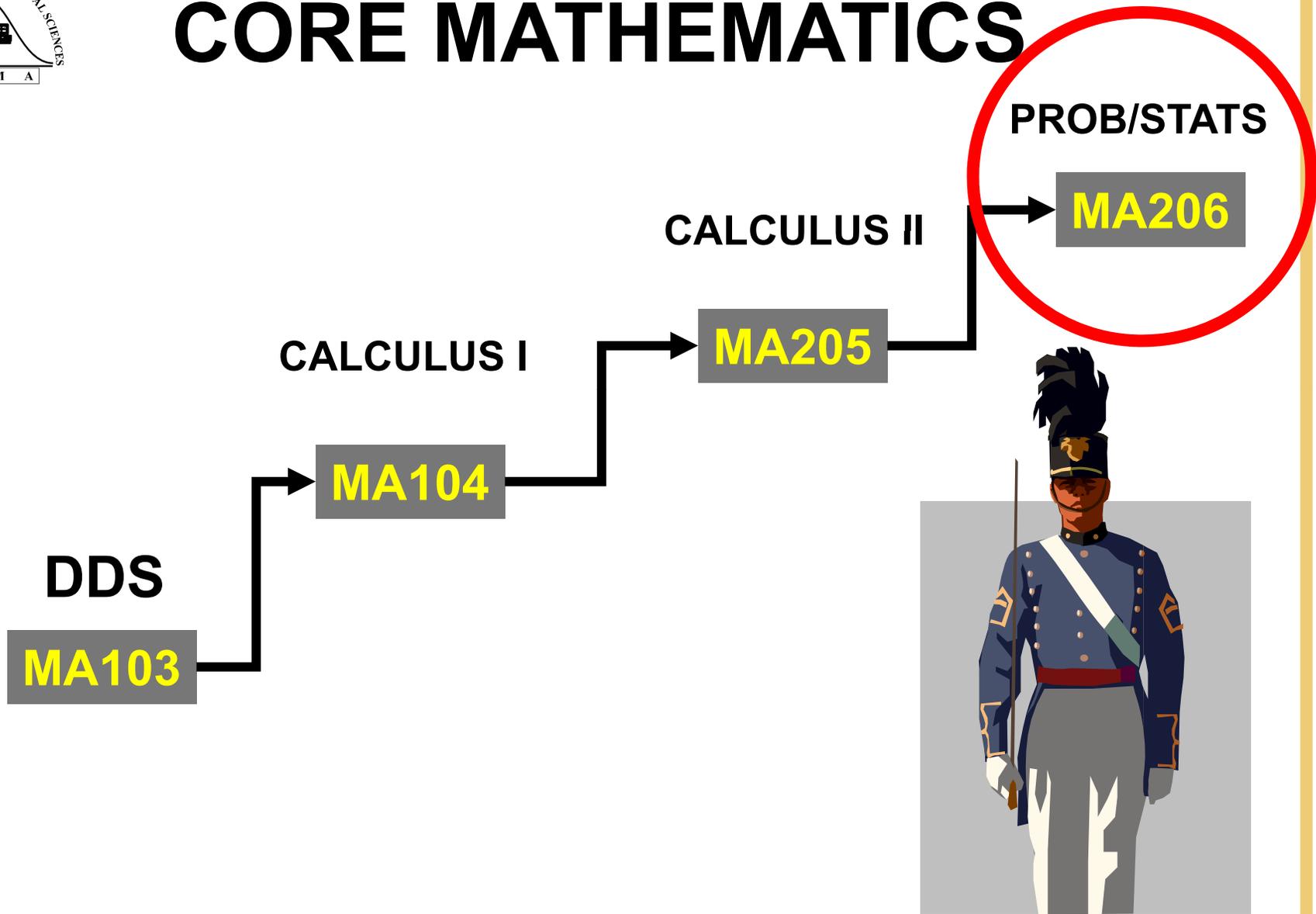


INTRODUCTION

- **LTC J. Scott Billie**
- **USMA - 1986**
- **Armor Branch/FA 49**
- **TH 253**
- **938-0209 (W)**
- **839-0036 (H)**
- **Teach A/C/G Hours**
- **john.billie@usma.edu**



CORE MATHEMATICS





EVALUATION PLAN

- **Course-wide events → course guide**
- **Grading scale → course guide**
- **Instructor Points**



CHAPTER 8

ACADEMIC PROCEDURES

800. General Academic Policies.

1. Personal Responsibilities. The primary responsibility for achieving satisfactory academic performance **rests with the individual cadet**. Cadets must develop a sense of personal responsibility for the achievement of the highest level of academic

achievement of which they are capable. To meet this responsibility, cadets have an obligation to know their academic status, manage their time, and establish effective priorities. The performance of academic duties is a significant part of the process of preparing for the acceptance of the duties and responsibilities of Army officers. The standard for performance of academic duties is the same as that for the performance of officer duties -- excellence and one's personal best.

USCC SOP
June 2008



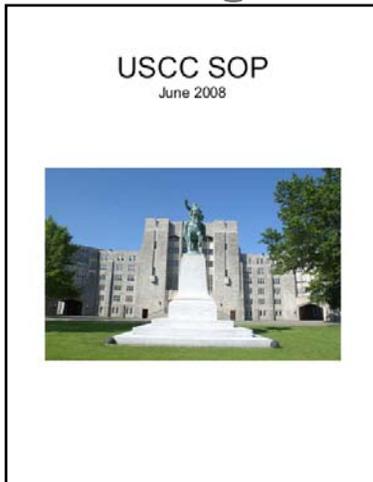


CHAPTER 8

ACADEMIC PROCEDURES

800. General Academic Policies.

5. Course Materials. Cadets will ensure that they **possess all required textbooks, course material and accessories prior to the start of each academic course**, except for any material designated for issue in class. Cadets will not rely on using others' materials; they must have sole use of the book or other item for the entire term. Cadets should also have an individual copy of the 1 day/2 day schedule (Buff Card) for the current academic year.



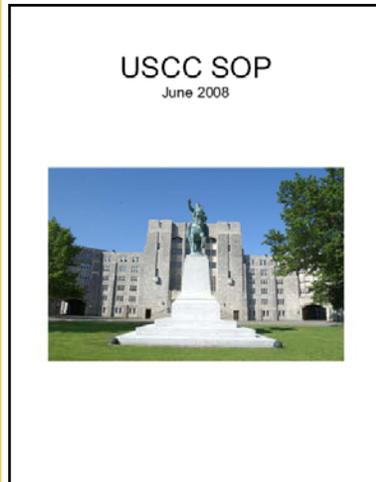


CHAPTER 8

ACADEMIC PROCEDURES

801. General Class Attendance Guidance.

3. Conflicts. Conflicts among required duties may arise occasionally. **Cadets are responsible to identify such conflicts as early as possible and resolve them expeditiously.**





CHAPTER 8

ACADEMIC PROCEDURES

802. Classroom Procedures.

2. Conduct.

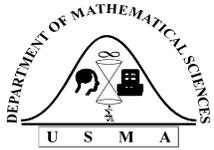
(c) Smoking, using smokeless tobacco, chewing gum, and consuming food is prohibited at all times in classrooms, lecture halls, and auditoriums.

(d) At Department Head discretion, drinks are permitted in selected classrooms during duty hours. All beverages, except plain water must be contained in an approved closed container. The only approved models of closed containers will be sold at the C-Store* and no other closed containers are permitted.

USCC SOP
June 2008



*For a nominal fee.



CHAPTER 8

ACADEMIC PROCEDURES

805. Additional Instruction (AI).

1. a. Academic.

(1) Additional instruction is instruction on course material supplemental to that presented during normal class time as scheduled in the course syllabus.

(2) Additional instruction is normally presented by the cadet's regularly assigned instructor, but in any case by an instructor of the department presenting the course.

(3) Material covered in AI is restricted to that previously covered during regularly scheduled class attendance. Coverage of future material (pre-teaching) is not permitted.

USCC SOP
June 2008





CHAPTER 8 ACADEMIC PROCEDURES

813. Additional Instruction (AI).

a. Academic.

(4) AI may be scheduled in accordance with the Academy schedule. Departments are encouraged to arrange times for AI that are mutually agreeable to cadets and the department, particularly in the cases of cadets who bear heavy institutional responsibilities. AI will not be conducted during the evening study period without the prior approval of the Dean. Although AI is not normally presented during the TEE period, specific questions addressed by cadets to instructors will be answered.

(5) Though AI is optional, once cadets have made appointments to receive AI or otherwise committed to attend AI at a specified time, attendance becomes mandatory and cadets will be reported if absent. Cadets are responsible to avoid conflicts or to resolve them should they occur.

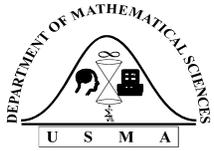
USCC SOP
June 2008





CLASS PROCEDURES

- **BEFORE CLASS**
 - Required materials
 - Appearance / conduct
- **FLOW OF CLASS**
 - Required materials
 - Questions from readings on boards
 - Section marcher gives report
 - Boards & briefing, quizzes
 - Lesson Link
 - **Use of Internet, IM, email in class**
- **ATTENDANCE**
 - Notification of Known Absence
 - Duty
 - WPR Attendance

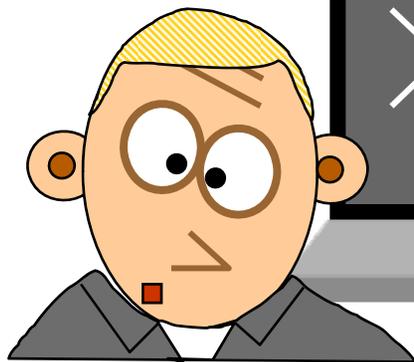
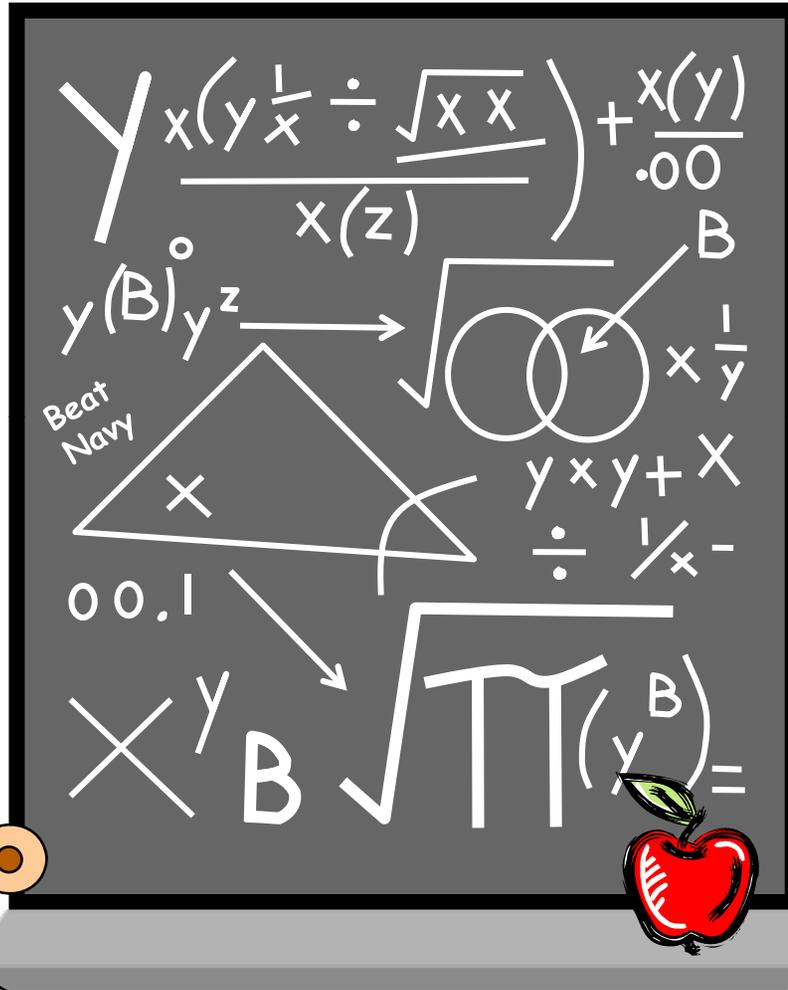


Admin Information

DOCUMENTATION OF WRITTEN WORK



Incorrect Blackboard Set up





Correct Blackboard Set up

1

$$f(x) = 5x^2 + x + 6$$

$$\underline{\underline{f'(x) = 10x + 1}} \text{ANS}$$

2

$$y = x^{-2/5}$$

$$\underline{\underline{y = -2/5 x^{-7/5}}} \text{ANS}$$

3

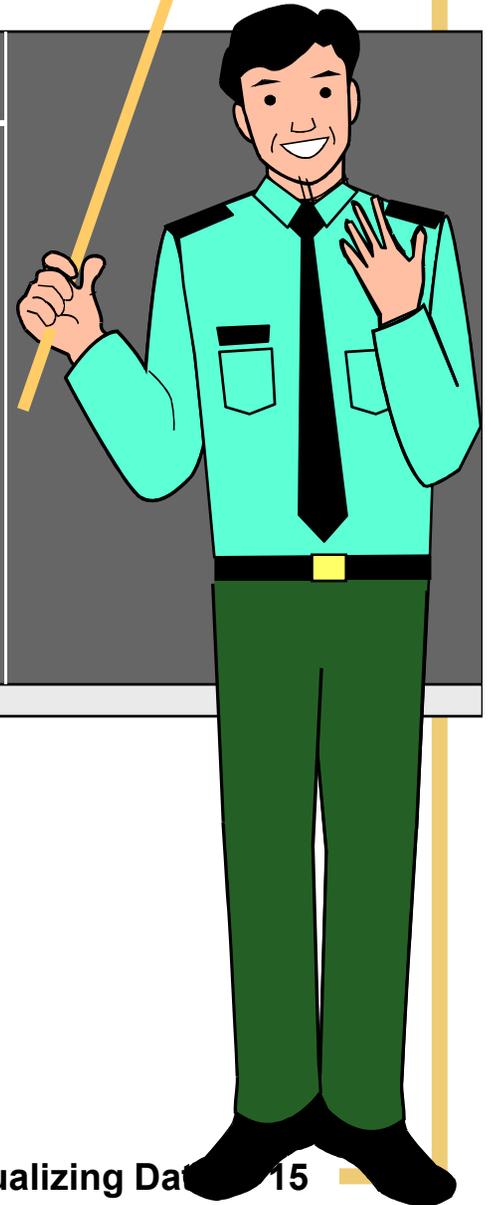
LTC Billie

$$h(x) = (x^2 + 4)(x^3 + 5)$$

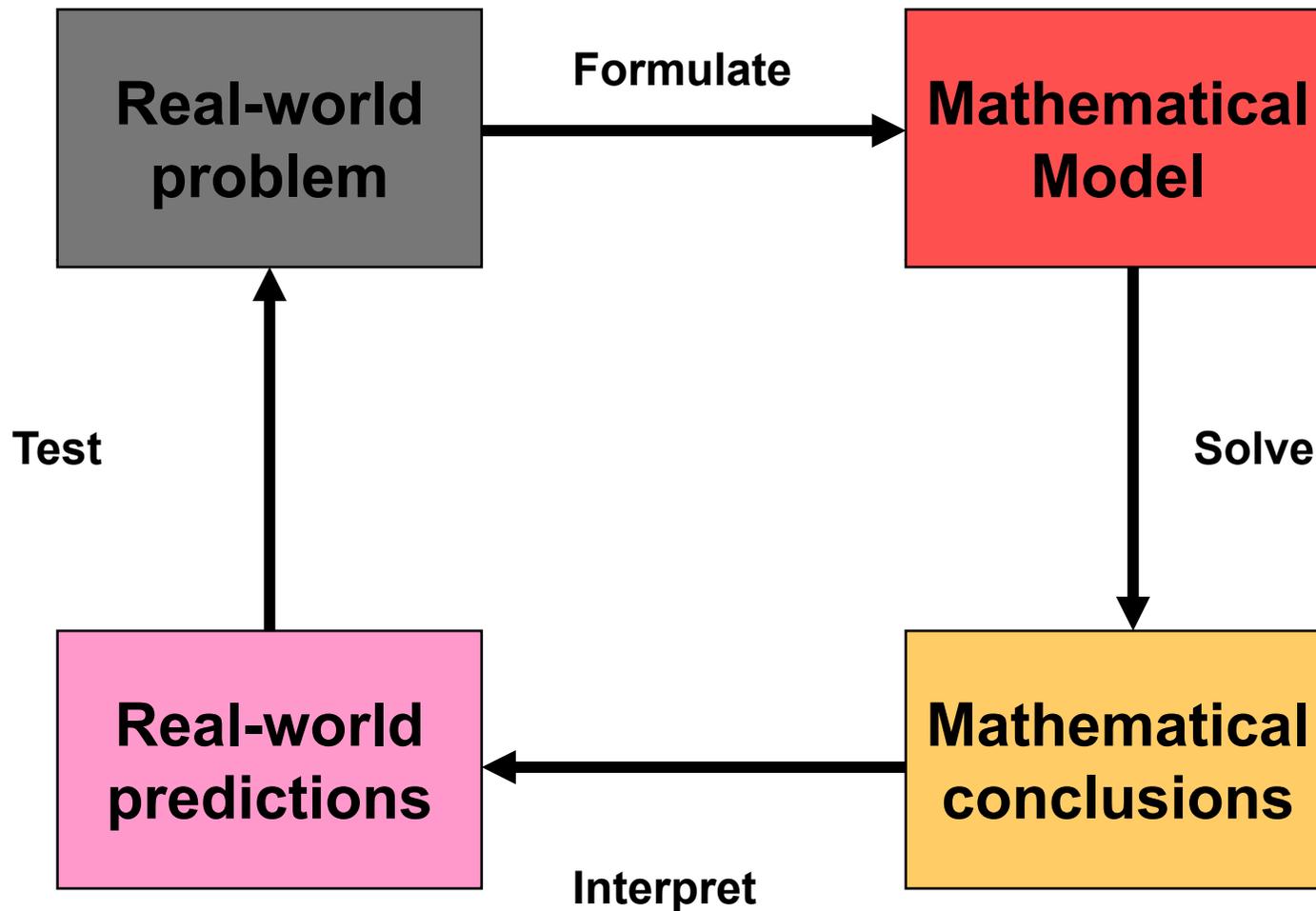
$$h'(x) = (x^2 + 4)3x^2 + (x^3 + 5)2x$$

$$= 3x^4 + 12x^2 + 2x^4 + 10x$$

$$\underline{\underline{= 5x^4 + 12x^2 + 10x}} \text{ANS}$$



Modeling Process





Objectives

Visualizing & Quantifying Data

- 1. Understand, construct, and interpret visual representations of data. (Histogram)
- 2. Calculate, interpret, and understand measures of location. (Mean, Median, Trimmed Mean, Percentiles.)
- 3. Calculate, interpret, and understand measures of variability. (Variance and Standard Deviation.)



Typical Data Set

23	60	79	32	57	74	52	70	82	36
80	77	81	95	41	65	92	85	55	76
52	10	64	75	78	25	80	98	81	67
41	71	83	54	64	72	88	62	74	43
60	78	89	76	84	48	84	90	15	79
34	67	17	82	69	74	63	80	85	61

What can you tell me
about those numbers?

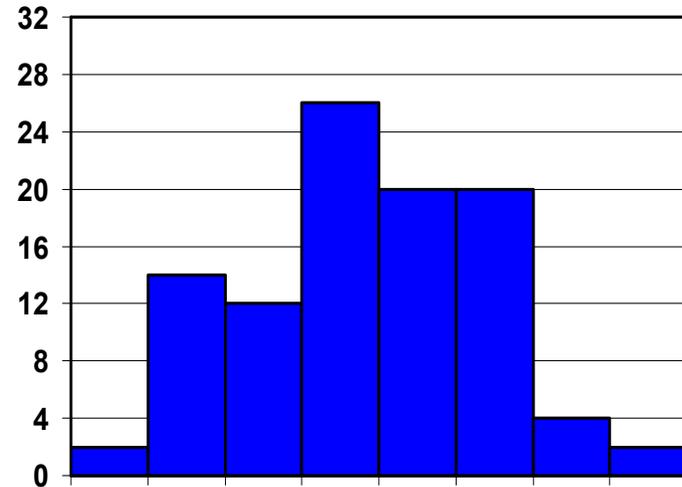
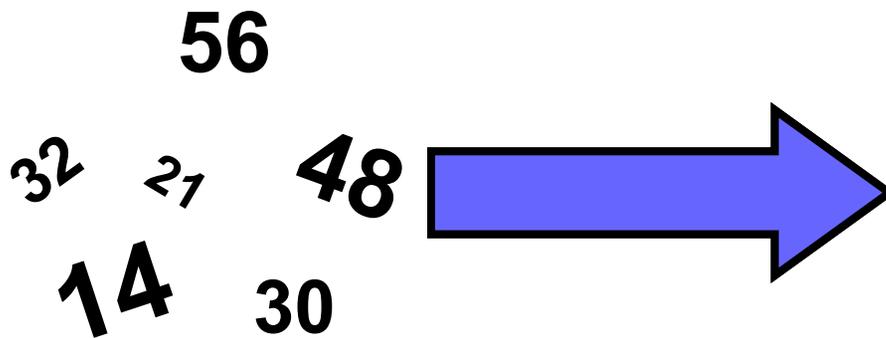




Descriptive Statistics

Definition

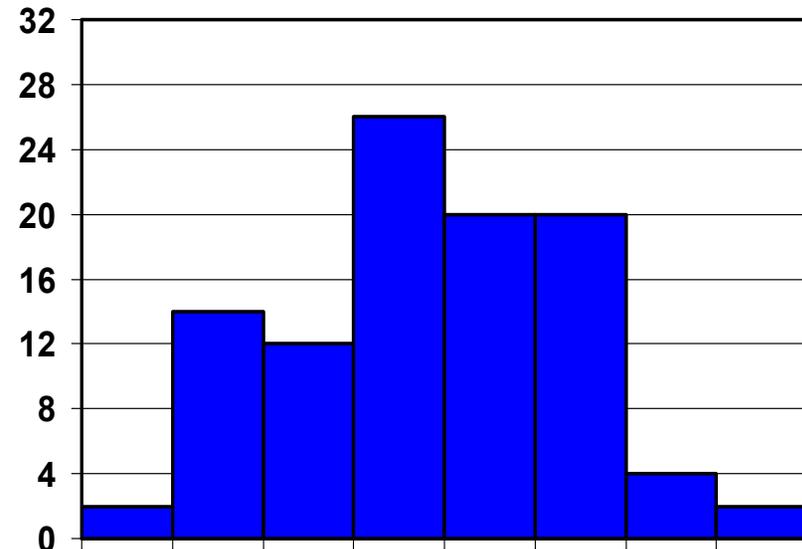
The reduction of data into into numerical and graphical summaries.

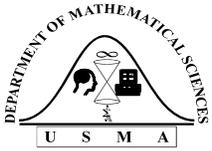




Graphical Representations of Quantitative Data

- Array
- Dot Plot
- Stem-and-Leaf Display
- Histogram



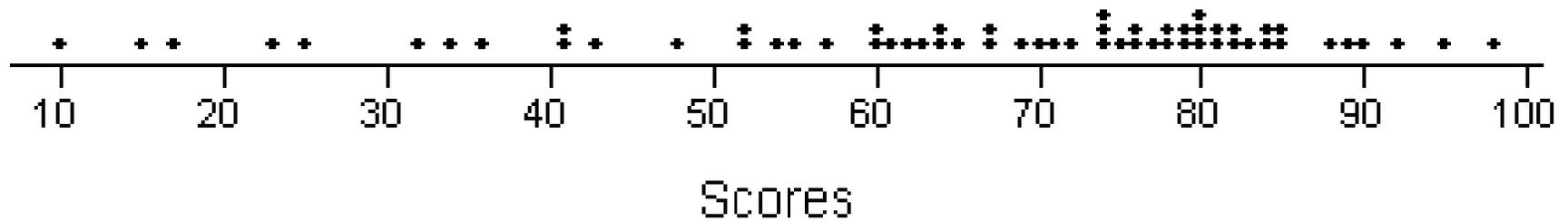


Array of Previous Data

10	43	62	72	79	84
15	48	63	74	79	84
17	52	64	74	80	85
23	52	64	74	80	85
25	54	65	75	80	88
32	55	67	76	81	89
34	57	67	76	81	90
36	60	69	77	82	92
41	60	70	78	82	95
41	61	71	78	83	98



Dot Plot (Minitab Example)



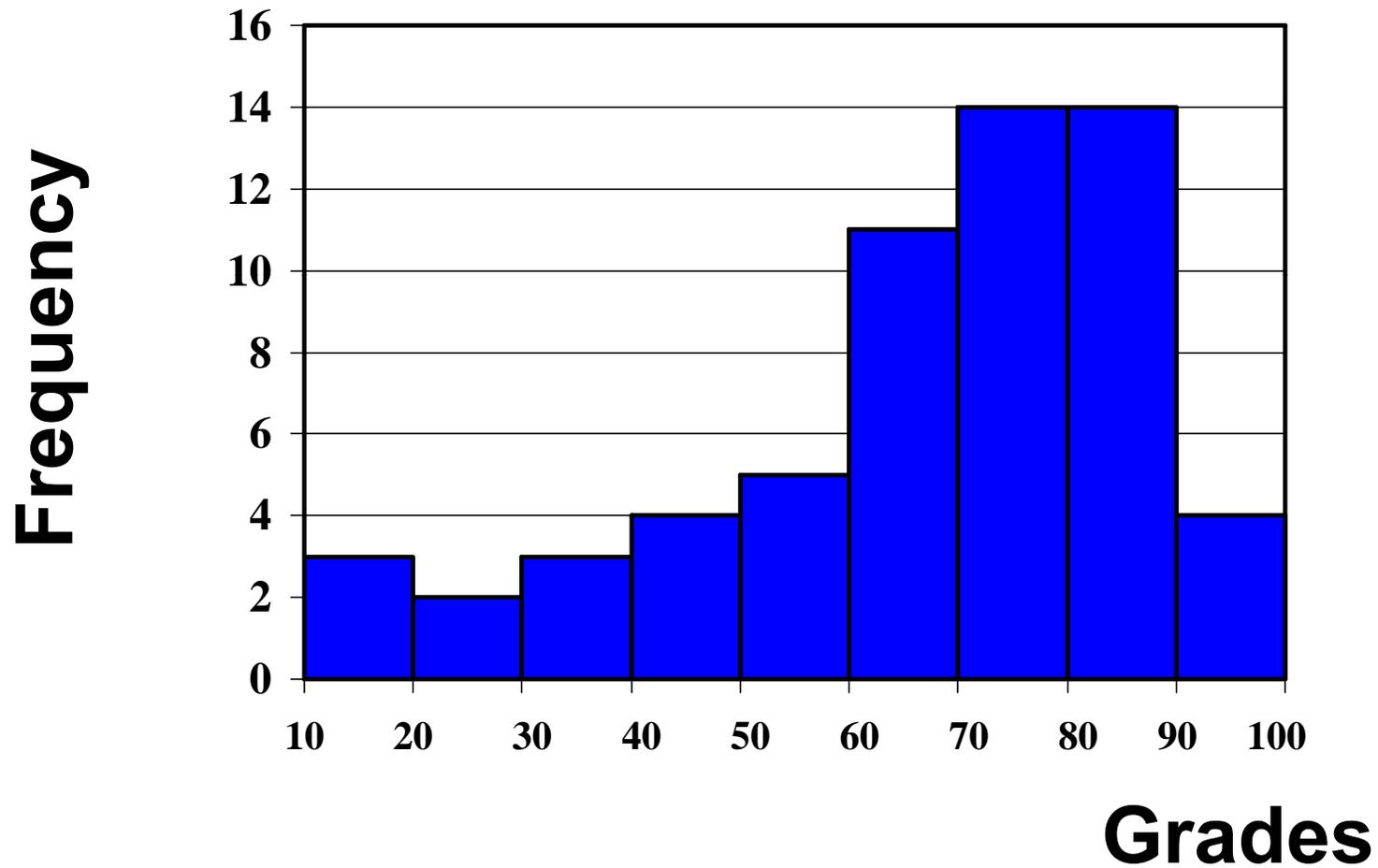


Stem-and-Leaf Display

Stem	Leaf	Frequency
1*	0	1
1.	5 7	2
2*	3	1
2.	5	1
3*	2 4	2
3.	6	1
4*	1 1 3	3
4.	8	1
5*	2 2 4	3
5.	5 7	2
6*	0 0 1 2 3 4 4	7
6.	5 7 7 9	4
7*	0 1 2 4 4 4	6
7.	5 6 6 7 8 8 9 9	8
8*	0 0 0 1 1 2 2 3 4 4	10
8.	5 5 8 9	4
9*	0 2	2
9.	5 8	2
		<hr/> 60



Histogram

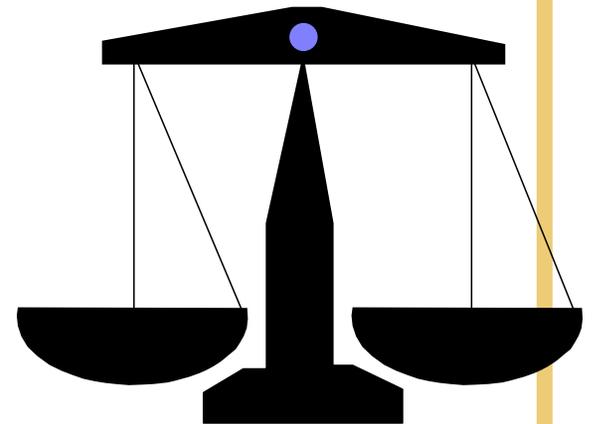




Mean

For a population: $\mu = \frac{\sum_{i=1}^N x_i}{N}$

For a sample: $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$





A sample of the math instructors at West Point was taken. The participants were asked to record their number of dependents.

The results are as follows:

1	2	2
1	2	5
1	2	





Median

For a population: $\tilde{\mu}$

For a sample: \tilde{x}





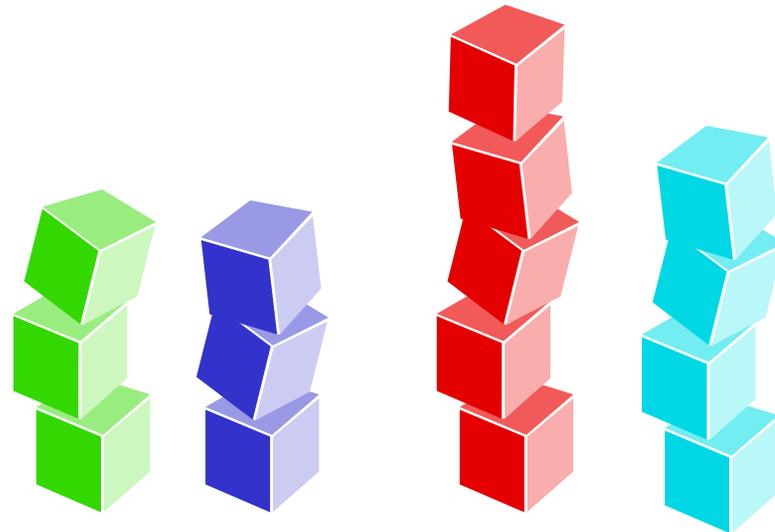
Median

To obtain the median for a data set, do the following:

- 1. Array data in ascending order.**
- 2. If the number of data points is odd, then use the middle observation.**
- 3. If the number of data points is even, then use the middle two data points and average.**

Mode

The mode, M , of a set of measurements is the value that occurs with the greatest frequency.

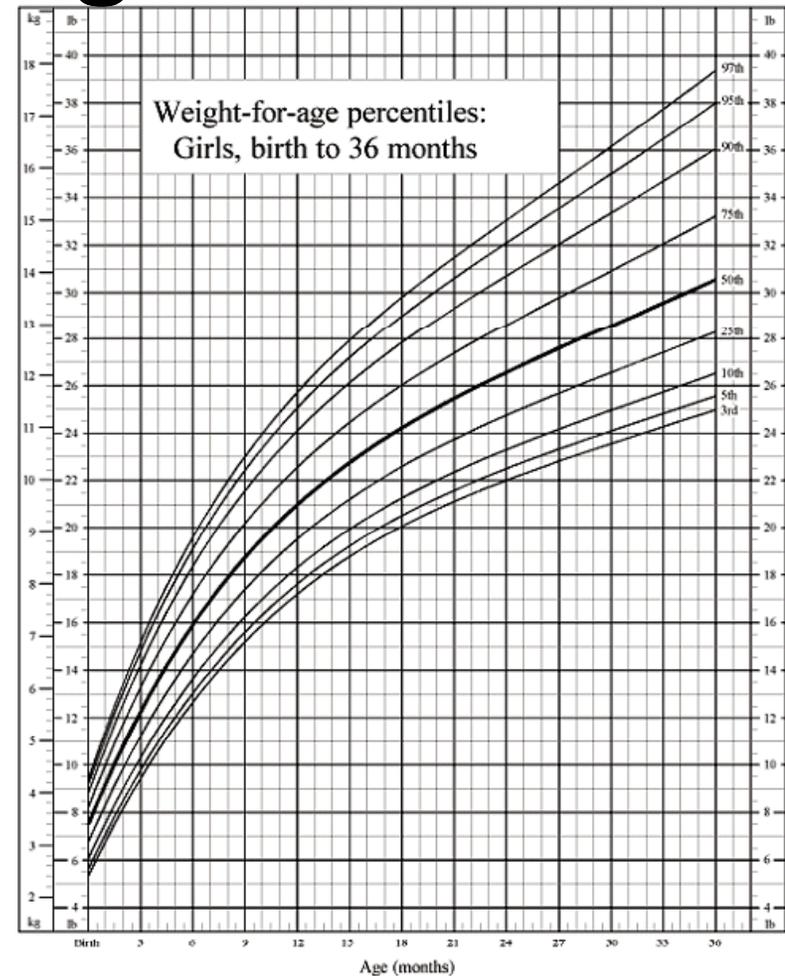




A Measure of Relative Standing

The p^{th} percentile is defined as a score such that p percent of the scores have numerical value less than or equal to that score.

CDC Growth Charts: United States



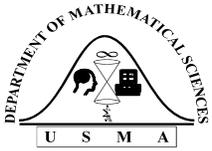
SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).





Percentile Definitions

- The 25th percentile is also called the 1st quartile.
- The 50th percentile is also called the 2nd quartile and the median.
- The 75th percentile is also called the 3rd quartile.



Interquartile Ranges

The interquartile range is the third quartile minus the first quartile.

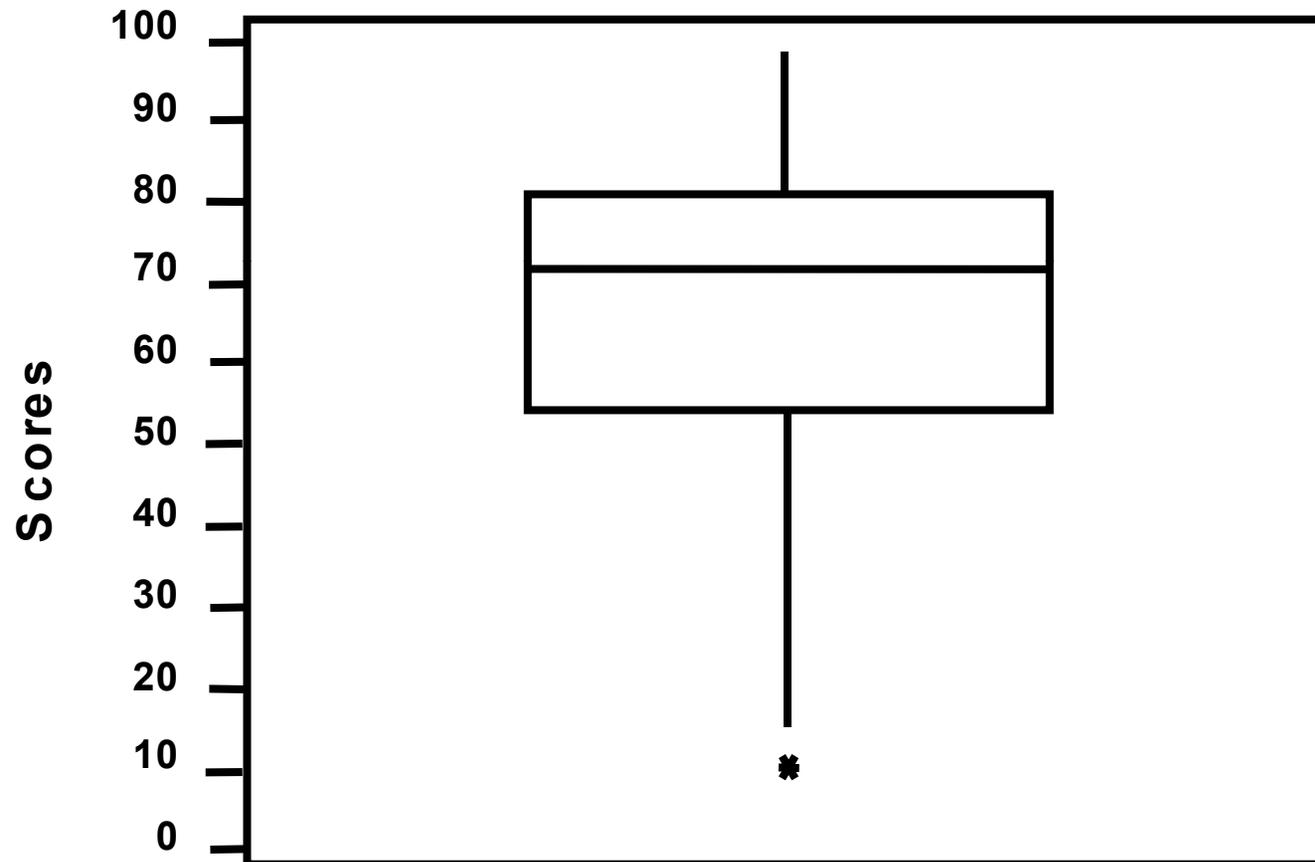
$$Q_3 - Q_1$$

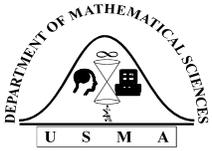
The semi-interquartile range is the third quartile minus the first quartile divided by two.

$$Q = \frac{Q_3 - Q_1}{2}$$



Box Plot

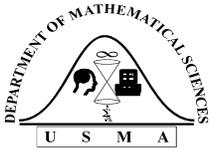




Outlier

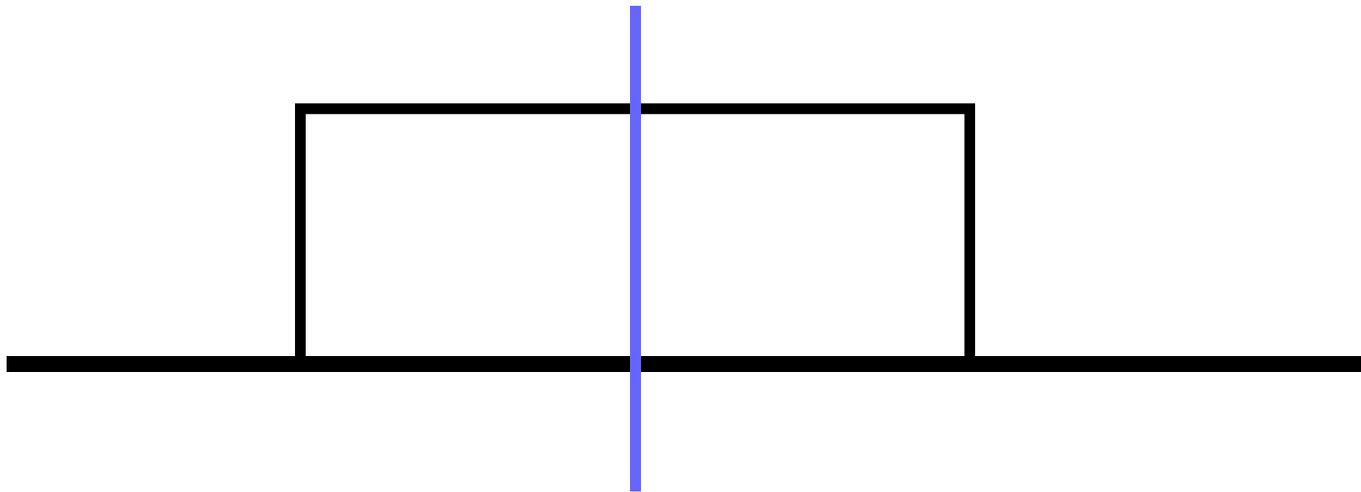
Outliers are observations that are considered to be extremely far removed from the bulk of the data. They are a rare event!

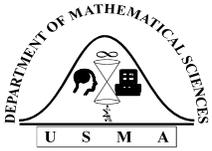
If the distance from the observation to the box exceeds 1.5 times the interquartile range (in either direction), then the observation may be labeled an outlier.



Symmetric

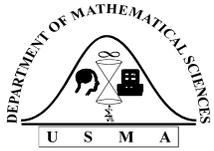
A distribution is said to be [symmetric](#) if it can be folded along a vertical axis so that the two sides coincide.



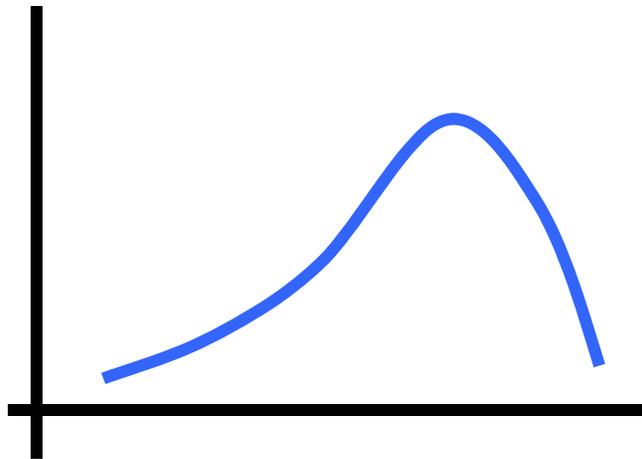


Skewed

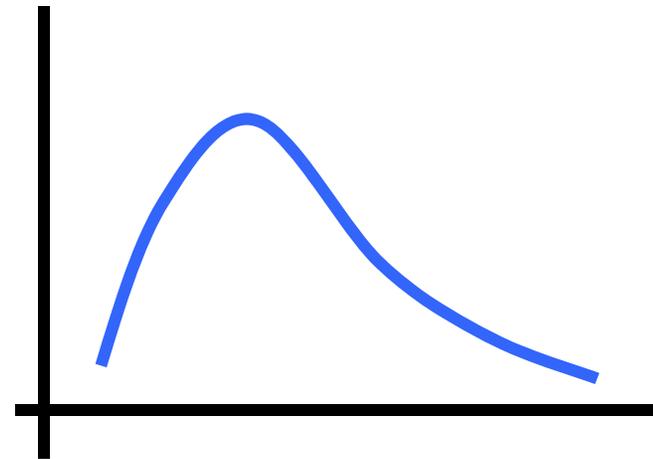
- If a distribution is not symmetric, it is skewed.
- If there are a few extremely low values, the distribution is said to be skewed left or negatively skewed.
- If there are a few extremely high values, the distribution is said to be skewed right or positively skewed.



Examples of Skewed Data



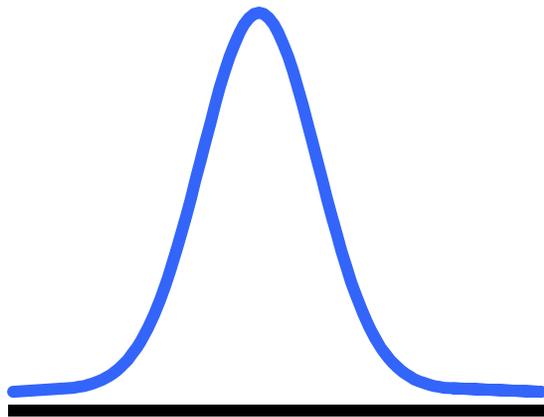
LEFT



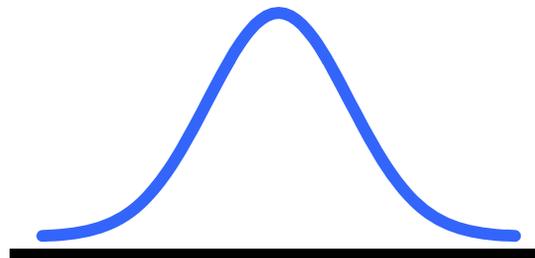
RIGHT



Kurtosis



Leptokurtic



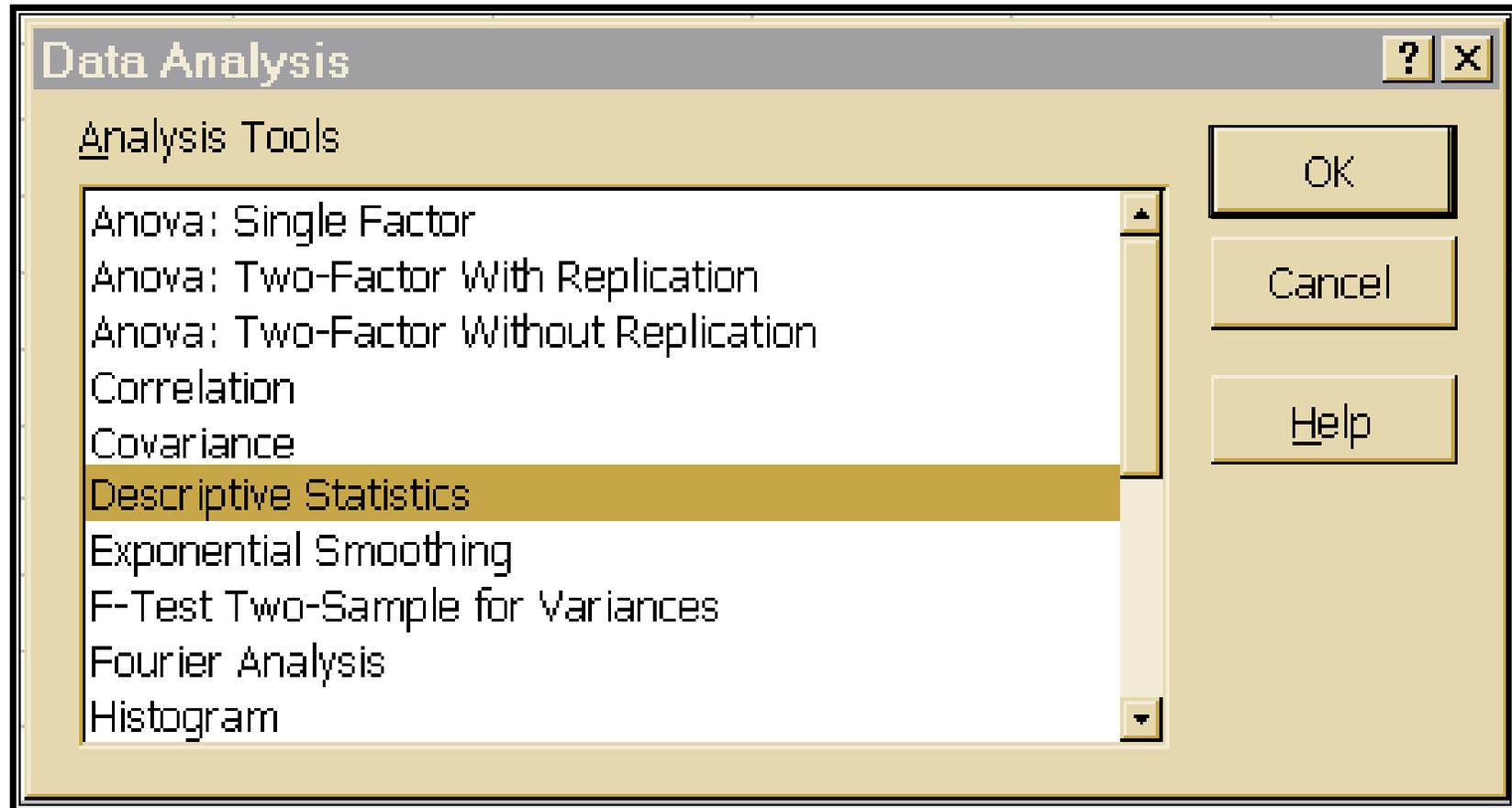
Mesokurtic



Platykurtic



Using Excel Spreadsheets





Excel Results

	A	B
1	Column1	
2		
3	Mean	2
4	Standard Error	0.46291005
5	Median	2
6	Mode	2
7	Standard Deviation	1.309307341
8	Sample Variance	1.714285714
9	Kurtosis	4.9
10	Skewness	2.036700309
11	Range	4
12	Minimum	1
13	Maximum	5
14	Sum	16
15	Count	8



Mathematica

Statistics`DescriptiveStatistics`

[Built-in Functions](#)
[Add-ons & Links](#)
[The Mathematica Book](#)
[Front End](#)
[Getting Started](#)
[Tour](#)
[Demos](#)
[Master Index](#)

Wolfram Research Products
 Standard Packages
 Combinatorica
 AuthorTools
 XML Capabilities
 MathLink
 J/Link

Geometry
 Graphics
 LinearAlgebra
 Miscellaneous
 NumberTheory
 NumericalMath
Statistics
 Utilities

DataManipulation
 DataSmoothing
DescriptiveStatistics
 DiscreteDistributions
 HypothesisTests
 LinearRegression
 MultiDescriptiveStatistics
 MultiDiscreteDistributions

Statistics`DescriptiveStatistics`

Descriptive statistics refers to properties of distributions, such as location, dispersion, and shape. The functions in this package compute descriptive statistics of lists of data. You can calculate some of the standard descriptive statistics for various known distributions by using the `Statistics`ContinuousDistributions`` and `Statistics`DiscreteDistributions`` packages. This package also provides some commonly used data transformations.

Note that this package is automatically loaded when most other statistical packages are used. For example, all the functions described below are available for use with the package `Statistics`HypothesisTests``.

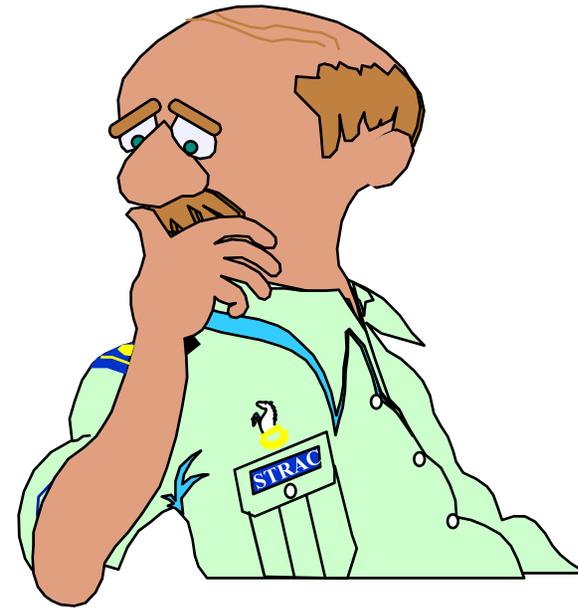
The statistics are calculated assuming that each value of data x_i has probability equal to $\frac{1}{n}$, where n is the number of elements in the data.

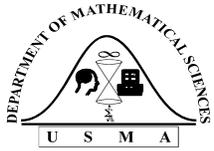
<code>Mean[data]</code>	average value $\frac{1}{n} \sum_i x_i$
<code>Median[data]</code>	median (central value)
<code>Mode[data]</code>	mode
<code>GeometricMean[data]</code>	geometric mean $(\prod_i x_i)^{\frac{1}{n}}$
<code>HarmonicMean[data]</code>	harmonic mean $n / \sum_i \frac{1}{x_i}$
<code>RootMeanSquare[data]</code>	root mean square $\sqrt{\frac{1}{n} \sum_i x_i^2}$
<code>TrimmedMean[data, f]</code>	mean of remaining entries, when a fraction f is removed from each end of the sorted list of data
<code>TrimmedMean[data, { f₁, f₂ }]</code>	mean of remaining entries, when fractions f_1 and f_2 are dropped from each end of the sorted data
<code>Quantile[data, q]</code>	q^{th} quantile
<code>InterpolatedQuantile[data, q]</code>	q^{th} quantile of the distribution inferred by linear interpolation of the entries in the list of data
<code>Quartiles[data]</code>	list of quartiles
<code>LocationReport[data]</code>	list of location statistics including <code>Mean</code> , <code>HarmonicMean</code> , and <code>Median</code>



Find the Mean, Median, & Mode

<u>X</u>	<u>Y</u>
1	49
50	50
50	50
99	51





Measures of Dispersion

- 1. Range**
- 2. Variance**
- 3. Standard deviation**
- 4. Interquartile range**

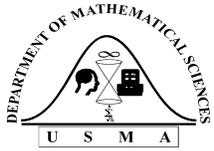


Range

The range is the difference between the largest and the smallest measurement in a data set.



$$\text{Sample Range} = X_{\max} - X_{\min}$$



Variance

For a population:

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



Variance

For a sample:

Derivational:

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

Computational:

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



Standard Deviation

The standard deviation is the positive square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

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