Duxbury Thomson Learning

Making Hard Decision

Third Edition

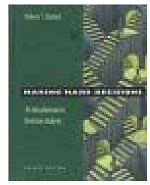
CHAPTER

7a

Modeling Uncertainty

A. J. Clark School of Engineering •Department of Civil and Environmental Engineering

FALL 2003



By

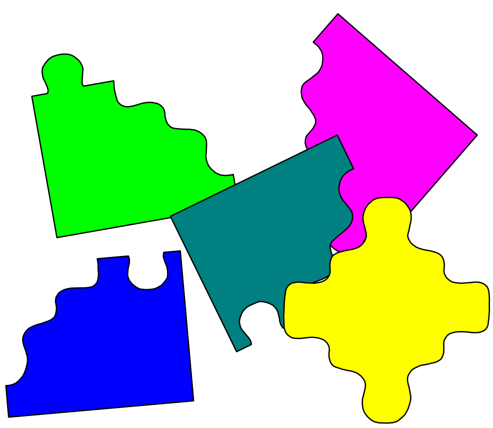
Dr . Ibrahim. Assakkaf

ENCE 627 – Decision Analysis for Engineering

Department of Civil and Environmental Engineering

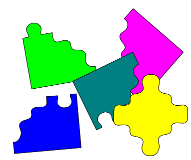
University of Maryland, College Park

Modeling Uncertainty



Modeling Uncertainty

- Uncertainty is a critical element of many decisions that we face.
- We will consider a variety of ways to model uncertainty in decision problems by using probability.
- To show some ways that probability modeling can be useful in decision problems.
- Give you a chance to polish your ability to manipulate probabilities.



Modeling Uncertainty

- Definitions
  - **Probability:** is a numerical measure of the likelihood of occurrence of an event relative to a set of alternative events (Ang and Tang 1975).
  - **Statistics:** is the estimation of certain parameters (i.e., mean,COV, distribution type) needed to quantify uncertainty and to describe the probability density function.

Modeling Uncertainty

– **Reliability:** is the probability of successful performance of an engineering system. It is the converse of the term "*probability of failure*". Reliability or risk assessment of an engineering system uses the methods of probability and statistics.

Probability + Statistics => Reliability

Modeling Uncertainty

- Decision Making in Engineering
  - Best Decision?
    - Full understanding of alternative solution procedures
      - Unbiased Solution
      - Highly precise
      - Cost effective
      - Have minimal environmental consequences

Modeling Uncertainty

- Decision Making in Engineering (cont'd)
  - Typical Approach to an Engineering Solution
    - Identify the problem
    - State the objective
    - Develop alternative solutions
    - Evaluate the alternatives, and
    - Use the best alternative

Modeling Uncertainty

- Decision Making in Engineering (cont'd)
  - Knowledge of probability, statistics, and reliability can help the engineer to ensure that each of the previously noted tasks are properly handled.
    - Identify the uncertainties associated with the alternative solutions.
    - Assess the array of possible outcomes of each alternative with their associated uncertainties.
    - Evaluate all data and prediction models used in the analyses, also with their associated uncertainties

Modeling Uncertainty

- Design of Engineering Systems
  - Design of engineering systems is usually a trade-off between maximizing safety and minimizing cost.
  - A design procedure that can accomplish both of these objective is highly desirable, but also difficult.

Modeling Uncertainty

- Design of Engineering Systems (cont'd)
  - Deterministic design procedures (i.e., ASD or WSD) do not provide adequate information to achieve the optimal use of the available resources to maximize safety and minimize cost.
  - On the other hand, probabilistic-based design can provide the required information for optimum design.
  - Probability, statistics, and reliability tools can help achieving the optimal design.

Modeling Uncertainty

- Need for Probability and Reliability Evaluation
  - The presence of ***uncertainty*** in engineering design and analysis has always been recognized.
  - Traditional approaches simplify the problem by considering the uncertain parameters to be deterministic.

Modeling Uncertainty

- Need for Probability and Reliability Evaluation
  - Traditional approaches account for the uncertainty through the use of empirical safety factor.
  - This factor is based on past experience but does not absolutely guarantee safety or performance.

Modeling Uncertainty

- Probability Based-design Approach Versus Deterministic Approach

$$\frac{R_n}{\mathrm{FS}} \geq \sum_{i=1}^{m} L_i$$

ASD

$$\phi R_n \geq \sum_{i=1}^{m} \gamma_i L_i$$

LRFD

  - According to ASD, one factor of safety (FS) is used that accounts for the entire uncertainty in loads and strength.
  - According to LRFD (probability-based), different partial safety factors for the different load and strength types are used.

Modeling Uncertainty

- Uncertainty in Engineering
  - Deterministic Methods
    - They deal with uncertainty by using factors of safety, which are not necessarily rational.
  - Example: Structural Design
    - In structural design, loads are arithmetically combed without proper consideration of their respective uncertainty.
    - Intuitively, loads need to be combined in a weighted scheme that reflect the uncertainty in each load type.

Types of Uncertainty

- Uncertainties in engineering systems are considered to be mainly attributed to ambiguity and vagueness in defining the variables or parameters of the systems and their relations.
- Ambiguity is due to noncognitive or quantitative source, while vagueness is due to cognitive or qualitative source.

Types of Uncertainty

- Uncertainty in an engineering system stems from two sources:
  1. Noncognitive (quantitative)
     - Inherent randomness in all physical observation
       - Repeated measurements of the same physical quantity do not yield the same value due to environment, test procedure, instrument, and observer.
     - Statistical uncertainties due to the use of limited information to estimate the characteristic of these parameters.

Types of Uncertainty

- Model uncertainties that are due to simplifying assumptions in
  - Analytical and prediction models
  - Simplified methods, and
  - Idealized representations of real performance

2. Cognitive (qualitative)

- Definition of certain parameters such as structural performance (failure or survival), quality, deterioration, and skill and experience of construction workers.

Types of Uncertainty

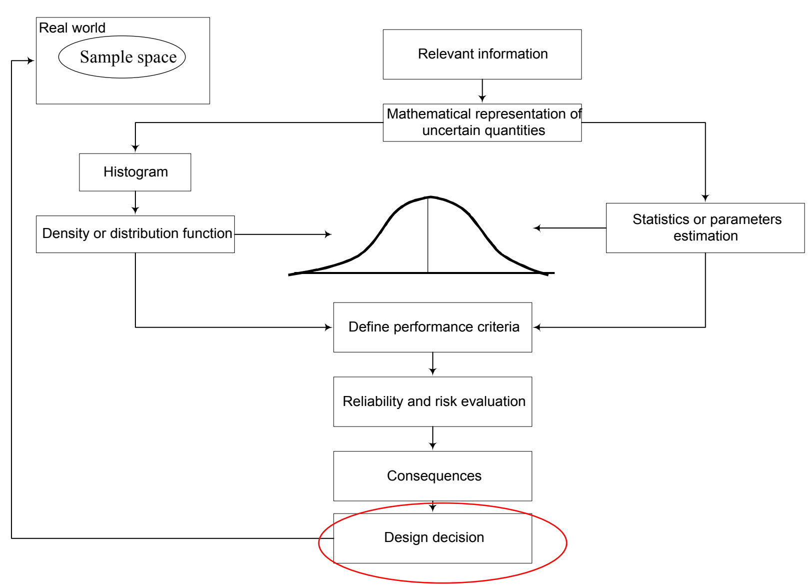
- Other human factors.
- Defining the interrelationships among the variables or parameters of the problems, especially for complex systems.

Note: Cognitive or qualitative source of uncertainty relates to the vagueness of the problem arising from intellectual abstraction of reality

Types of Uncertainty

- Example: Illustration of the three sources of noncognitive (quantitative) uncertainty
  - Suppose that the wind pressure acting on a building needs to be estimated (lb/ft^2). Record of wind speed data (mph) can be collected for the site. It is known that the wind speed cannot be predicted with certainty: thus, it is ***inherently random***. Its ***statistical uncertainty*** can be estimated by considering past observations, and more data lead to better estimate. However, the statistical information on wind speed needs to be converted to wind pressure, for which Bernoulli's principle is commonly used. This introduces another source of uncertainty, known as ***modeling uncertainty***.

Modeling of Uncertainty



Modeling of Uncertainty

- Example 1: chance event — finding hydrocarbons

Outcomes
1. hydrocarbon found
2. no hydrocarbon found (i.e. a dry hole)

- Example 2: chance event — Weather Condition

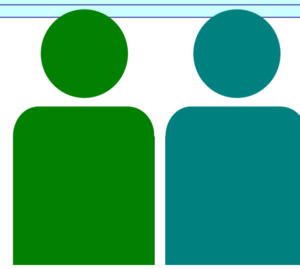
Outcomes
1. Rainy
2. Sunny

Modeling of Uncertainty

- Uncertainty issues can be related to the decision-maker's personal preferences. In other words his/her subjective intuition.
- Hence there is a relation between uncertainty and subjective probability.

Uncertainty & Subjective Probability

Most of us are comfortable making information statements that reflect our uncertainty. Example: We use terms such as "there's a chance that such-and-such will happen." In a decision-analysis approach, however, there is a need for more precision.



Uncertainty & Subjective Probability

- We can use probability to model subjective beliefs about uncertainty.
- Our beliefs and feelings about uncertainty can be translated into probability numbers that can be included in a decision tree or influence diagram.

Uncertainty & Subjective Probability

- In many cases the uncertainty that we face has characteristics that make it similar to certain prototypical situations.
- It may be possible to represent the uncertainty with a standard mathematical model and then derive probabilities on the basis of the mathematical model.

Uncertainty & Subjective Probability

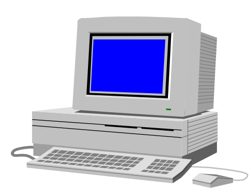
- We use historical data as a basis for developing probability distributions.
- If data about uncertainty in a decision situation are available, a decision maker would surely want to use them.
- How to use data alone to create histograms and continuous distributions.

Uncertainty & Subjective Probability

- We discuss the use of data to model relationships among variables
- We develop an approach that uses data to update a decision maker's probability beliefs via Bayes' theorem.
- "Create data" through computer simulation, or by what is known as Monte Carlo simulation.

Uncertainty & Subjective Probability

- One can construct a model of a complex decision situation and use a computer to simulate the situation many times.
- By tracking the outcomes, the decision maker can obtain a fair idea of the probabilities associated with various outcomes.



Uncertainty & Subjective Probability

- When faced with uncertainty most decision makers do their best to reduce it.
- The basic strategy that we follow is to collect information.

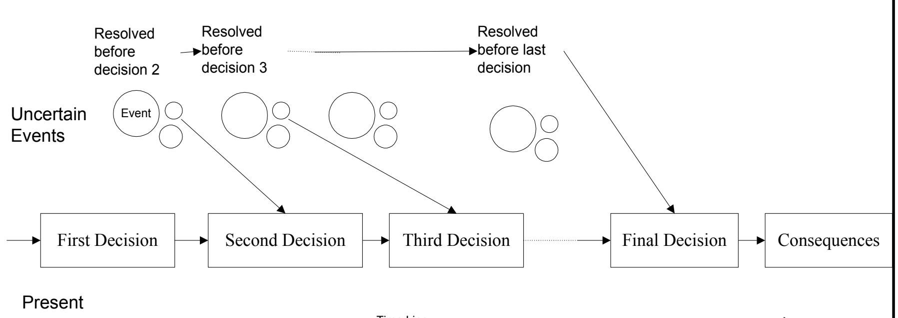
Uncertainty & Subjective Probability

- Determining what information is appropriate and then processing it can be costly. How much is information worth to you?
- A problem with many sources of uncertainty, calculating the value of information can help to guide the decision analysis, thus indicating where the decision maker can best expend resources to reduce uncertainty.

Error and Uncertainty

- Error:
  - A recognizable deficiency in any phase or activity of modeling and simulation that is not due to lack of knowledge
- Uncertainty:
  - A potential deficiency in any phase or activity of the modeling process that is due to the lack of knowledge.

Uncertainty and Decisions



Consequences

- For example failure consequences can include
  - Production loss including delays
  - Property damage that includes repair
  - Loss of life and injuries
  - Ecological effects
  - Various types of environmental damage
  - Social and cultural impacts

Sources of Uncertainty

- **Non-specificity (Ambiguity)**
  - Many engineering systems can have more than one outcome. The possibility of more than one outcome introduces this type of uncertainty (Hartley 1928). The likelihood of occurrence of each outcome is not a factor in defining this uncertainty type.

- **Example: Outcomes of Events**
  - The outcomes of the event "a naval vessel operated with a medium speed in a high sea state" are:
  - Structural damage (or failure)
  - No structural damage (or survival)

Structural damage and no structural damage need to be precisely defined.

Sources of Uncertainty

- **Probability (Ambiguity)**
  - The second source of uncertainty is related to the likelihood of occurrence (or probabilities) of outcomes (Shannon 1948).

- **Example: Likelihood of Damage**
  - In this, consider the following two vessels:
    - Vessel 1 with a likelihood of damage of 1 in 100 trials
    - Vessel 2 with a likelihood of damage of 1 in 5 trials

Vessel 2 can be considered to have more uncertainty than Vessel 1.

Sources of Uncertainty

- **Physical Randomness**
  - This uncertainty source can be attributed to inherent randomness in a parameter.

- **Example: Steel Shipment**

A steel shipment with a specified yield strength of 36 ksi was received. A sample was taken and tested. The following results were obtained:

Sample 1 (ksi)	**37.8**	**38.2**	**36.2**	**40.2**	**38.5**	**36.9**

Average = (37.8+38.2+36.2 +40.2 +38.5 + 36.9)/6 = 37.97 ksi
Different values were obtained from the different specimens.
This is the inherent variability in steel strength.

Sources of Uncertainty

- **Statistical Uncertainty**
  - Different samples can produce different results, hence this uncertainty source.
- **Example: Steel Shipment**

In this example, a second sample was taken with the following results:

Sample 1 (ksi)	**37.8**	**38.2**	**36.2**	**40.2**	**38.5**	**36.9**
Sample 2 (ksi)	**38.3**	**37.6**	**35.9**	**39.2**		

Average 1 = 37.97 ksi, Sample size = 6
Average 2 = 37.75 ksi, Sample size = 4
Different samples produce different results. The weighted average of the two averages is:
Average = [37.97(6)+37.75(4)]/10 = 37.88 ksi
Therefore, the average is also uncertain.

Sources of Uncertainty

- **Organizational and Human Errors**
  - The possibility of organizational and human errors introduces another source of uncertainty. Human errors can be, for example, due to ignorance, negligence, or miscommunication. Organizational errors can be, for example, as a result of ill defined procedures, miscommunication, loosely defined responsibility, bureaucracy, inefficiency, etc.

Example: Errors

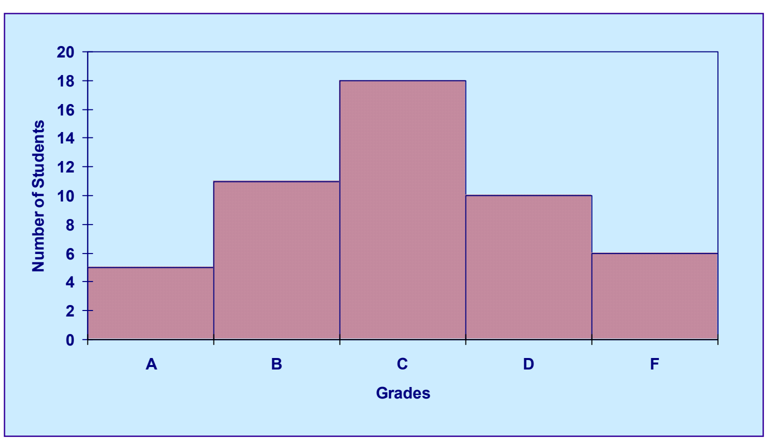
- Design errors
- Construction errors
- Limitations of specifications

Histograms and Frequency Diagrams

- Histogram
  - A histogram is a plot (or a tabulation) of the number of data points versus selected intervals or values for a parameter.
- Frequency Diagrams
  - A frequency diagram (or frequency histogram) is a plot (or a tabulation) of the frequency of occurrence versus selected intervals or values of the parameter.

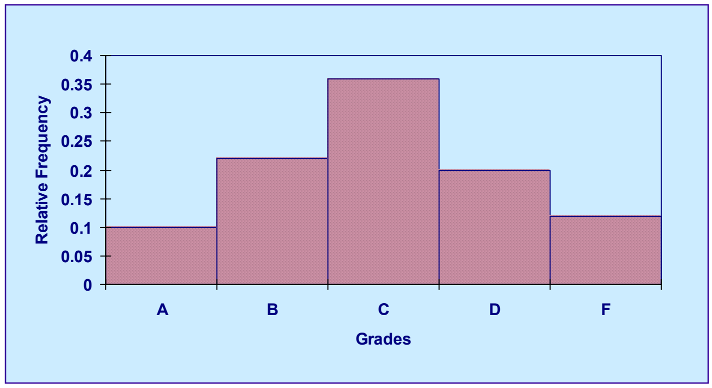
Histograms and Frequency Diagrams

- Example: Grade Histogram



Histograms and Frequency Diagrams

- Example: Grade Frequency Diagram



Histograms and Frequency Diagrams

- The number of intervals (k) can be subjectively selected depending on the sample size, n.
- The number of intervals can be approximately determined as

$$\boxed{k = 1 + 3.3\log_{10}(n)}$$

Histograms and Frequency Diagrams

- Also, the number of interval can depend on the level of dispersion in the data.
- The frequency diagrams (or histograms) can be derived from the histograms by dividing the number of data points that correspond to each interval by the sample size.

Histograms and Frequency Diagrams

- Example:

The starting salaries (in thousands of dollars) of 20 graduates, chosen at random from the graduating class of an urban university, were determined and recorded in the following table:

34	29	27	39	41
28	32	37	35	36
23	31	33	34	29
27	35	29	30	32

Draw a histogram and frequency diagram of the the graduate salaries.

Histograms and Frequency Diagrams

Sorted Data

23	32
27	33
27	34
28	34
29	35
29	35
29	36
30	37
31	39
32	41

Number of Data Points = 20, Min = 23, Max = 41

Data Range = 41 – 23 =18

$$k = 1 + 3.3\log_{10}(n) = 1 + 3.3\log_{10}(20) = 5.29$$

Take k = 5

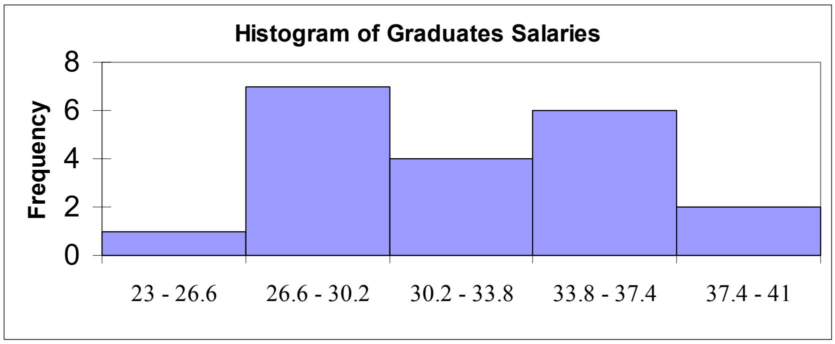
Hence, interval width = 18/5 = 3.6

The following histogram table and graphs can be constructed:

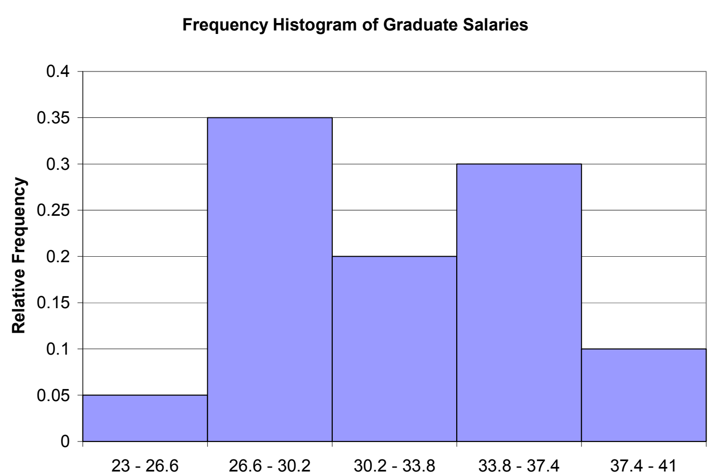
Histograms and Frequency Diagrams

Interval	Frequency	Relative Frequency
23.0 - 26.6	1	0.05
26.6 - 30.2	7	0.35
30.2 - 33.8	4	0.2
33.8 - 37.4	6	0.3
37.4 - 41.0	2	0.1
Total =	20	1

Histograms and Frequency Diagrams



Histograms and Frequency Diagrams



Histograms and Frequency Diagrams

Example: Rolling of a Pair of Dice

SECOND DIE

FIRST DIE

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

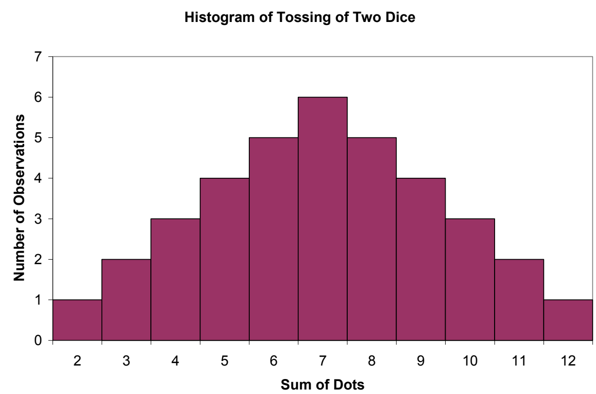
Histograms and Frequency Diagrams

- Example: Rolling of a Pair of Dice

  If a pair of dice rolled simultaneously, the histograms for the sum of dots from the two dice would appear as shown in the following tables and figures:

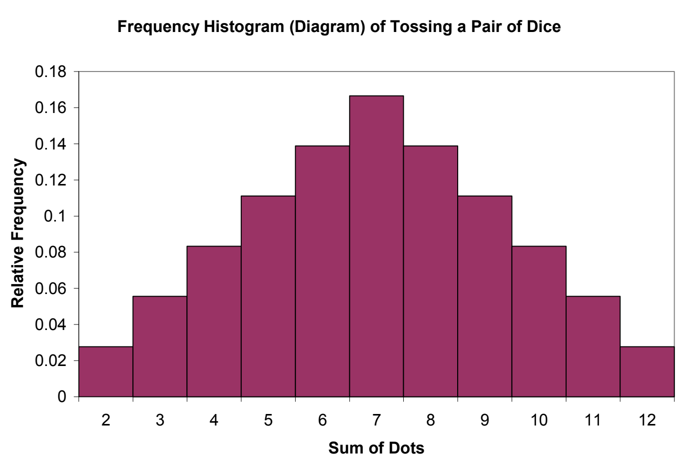
Histograms and Frequency Diagrams

Sum of Dots	No. of Observations
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1



Histograms and Frequency Diagrams

Sum of Dots	Relative Frequency
2	0.028
3	0.056
4	0.083
5	0.111
6	0.139
7	0.167
8	0.139
9	0.111
10	0.083
11	0.056
12	0.028



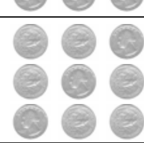
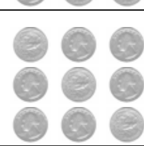
Histograms and Frequency Diagrams

- Example: Tossing of 3 Coins

  Suppose that we are interested in the number of heads (0, 1, 2, or 3) appearing on each toss of the three coins, then the outcome would be the following:

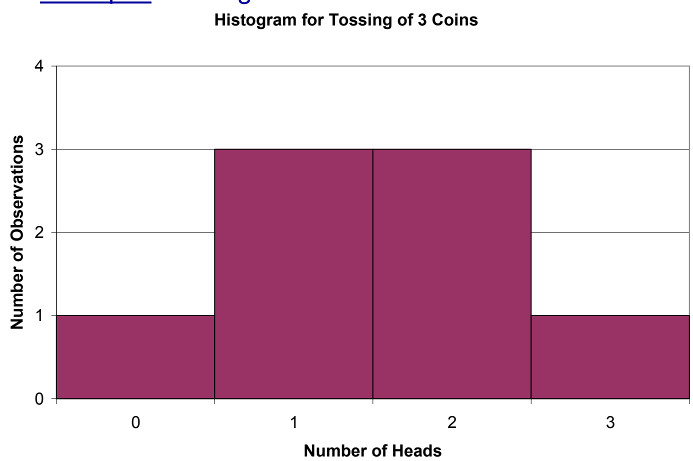
Histograms and Frequency Diagrams

Example: Tossing of 3 Coins

Outcome		Number of Heads	Frequency
TTT		0	1
(TTH), (THT), and (HTT)		1	3
(THH), (HTH), and (HHT)		2	3
(HHH)		3	1

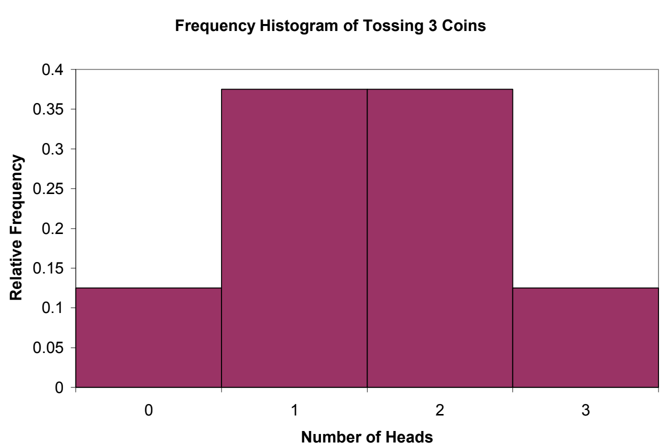
Histograms and Frequency Diagrams

Example: Tossing of 3 Coins



Histograms and Frequency Diagrams

Example: Tossing of 3 Coins



Descriptive Measures

- There are three fundamental types of measures for data analysis:
  1. Central Tendency Measures
     - Average Value
     - Median Value
     - Mode Value
  2. Dispersion Measures
  3. Percentile Measures

Central Tendency Measures

- These measures are very important descriptors of data.
- The following three types can be used:
  1. Average (Mean) Value
  2. Median Value
  3. Mode Value

Central Tendency Measures

- Average (Mean) Value
  - For n observations, if all observations are given equal weights, the average value can be given by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where x_i = a sample point, and i =1, 2, …,n

Central Tendency Measures

- Example: Average or Mean Value

Find the average (mean) for the sample measurements 3, 5, 1, 8, 6, 5, 4, and 6.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(3+5+1+8+6+5+4+6) = \frac{38}{8} = 4.75$$

Central Tendency Measures

- Median Value
  - The median value x_m is defined as the point that divides the data into two equal parts.
  - 50% of data are above x_m and 50% are below x_m.
  - The median value can be determined by ranking the *n* values in the sample in increasing or decreasing order.

Central Tendency Measures

- Steps for Computing the Median, x_m
  1. If *n* is an odd number, the median is the value with a rank of $(n + 1)/2$.
  2. If *n* is an even number, the median equals the average of the two middle values, that is, those with ranks $n/2$ and $(n/2) + 1$.

Central Tendency Measures

- Example: Median Value

Sorted Data

	Set 1	Set 2
	5	9
	7	10
	8	15
	13	20
	21	23
	-	24
$n =$	5	6

Find the the median for the following sets of measurements:

Set 1: 5, 21, 8, 7, and 13

Set 2: 10, 9, 23, 15, 20, and 34

$$\text{Median}_{\text{Set 1}} = 8$$

$$\text{Median}_{\text{Set 2}} = \frac{15 + 20}{2} = 17.5$$

Central Tendency Measures

- Mode Value
  - The mode value x_d is defined as the point of highest percent for the frequency of occurrence.
  - This point can be determined with the aid of Histogram or frequency histogram (diagram)

Central Tendency Measures

- Example 1: Mode Value

Number	Frequency	Relative Frequency
1	4	0.44
2	2	0.22
4	1	0.11
5	1	0.11
9	1	0.11
Total =	9	1

Find the mode for the following set of data:

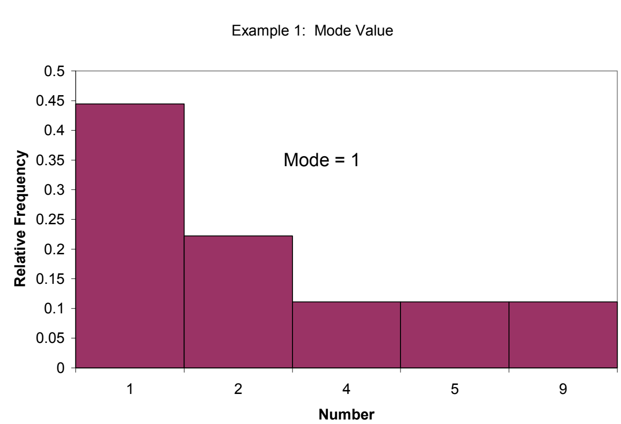
2, 1, 2, 1, 1, 5, 1, 9, and 4

The data can be arranged in ascending order as follows:

1, 1, 1, 1, 2, 2, 4, 5, 9

Hence, the mode = 1

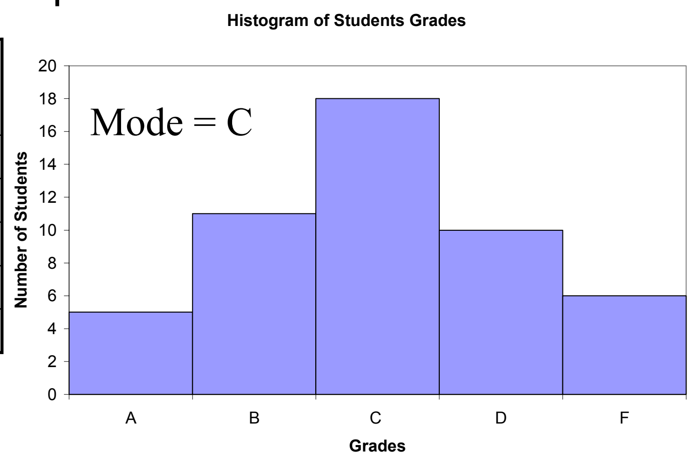
Central Tendency Measures



Central Tendency Measures

- Example 2: Mode Value

Grades	Number of Students
A	5
B	11
C	18
D	10
F	6



Dispersion Measures

- A measure of central tendency gives us a typical value that can be used to describe a whole set of data, but it <u>does not tell</u> us whether the data are tightly clustered or widely dispersed.
- Therefore, the dispersion measures describe the level in the data about the central tendency location

Dispersion Measures

- The dispersion measures include:
  - Range
  - Variance, S^2
  - Standard Deviation, S
  - Coefficient of Variation, COV or δ
  - Percentiles
  - Box-and-Whisker Plots

Dispersion Measures

- Range:

  "The range of a set of data is the difference between the largest and smallest (extreme values) values in the data set."

Example:

2.3, 1.2, 4.6, 10.4, 8.0

Therefore:

Range = Max – Min = 10.4 – 1.2 = 9.2

Dispersion Measures

- Variance, S^2
  - The sample variance of a set of n sample measurements $x_1, x_2, \ldots, x_n$ with mean $\overline{X}$ is given by

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2} \qquad (7\text{-}1)$$

NOTE: the factor (n – 1) is used instead of n to obtain an unbiased estimate of S^2

Dispersion Measures

- Example: Variance

  Find the variance for the following sample measurements:

  1, 3, 5, 4, and 3

$$n = 5$$

$$\overline{X} = \frac{1+3+5+4+3}{5} = 3.2$$

$$S^2 = \frac{(1-3.2)^2 + (3-3.2)^2 + (5-3.2)^2 + (4-3.2)^2 + (3-3.2)^2}{5-1} = 2.20$$

Dispersion Measures

– For computational purposes, the following alternative set of equations can be used to compute the variance:

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

OR

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{X}^2\right] \qquad (7\text{-}2)$$

Dispersion Measures

Example: Find the variance for the sample measurements 1, 3, 5, 4, and 3 using Eq. 2-2.

$$S^2 = \frac{1}{5-1}\left[1^2 + 3^2 + 5^2 + 4^2 + 3^2 - \frac{1}{5}(1+3+5+4+3)^2\right] = \frac{8.8}{4} = 2.2$$

OR

$$S^2 = \frac{1}{5-1}\left[1^2 + 3^2 + 5^2 + 4^2 + 3^2 - 5(3.2)^2\right] = \frac{8.8}{4} = 2.2$$

Dispersion Measures

- Standard Deviation, *S*

The standard deviation by definition is the square root of the variance. It is given by

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

OR

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{X}^2\right]} \qquad \text{(7-3)}$$

Dispersion Measures

- Coefficient of Variation, *COV* or δ

The coefficient of variation is a normalized quantity based on the standard deviation and the mean. It is a dimensionless quantity. The *COV* is defined as

$$COV = \frac{\text{standard deviation}}{\text{mean (or average)}} = \frac{S}{\overline{X}} \qquad \text{(7-4)}$$

Dispersion Measures

- Example: Dispersion Measures of Concrete Strength

A sample of five tests was taken to determine the compression strength (ksi) of concrete. Tests results are 2.5, 3.5, 2.2, 3.2, and 2.9 ksi. Compute the variance, standard deviation, and coefficient of variation of concrete strength.

Dispersion Measures

- Example (cont'd): Concrete Strength

$$\overline{X} = \frac{2.5 + 3.5 + 2.2 + 3.2 + 2.9}{5} = 2.86 \text{ ksi}$$

$$S^2 = \frac{2.5^2 + 3.5^2 + 2.2^2 + 3.2^2 + 2.9^2 - \frac{(2.5 + 3.5 + 2.2 + 3.2 + 2.9)^2}{5}}{5 - 1} = 0.273 \text{ ksi}^2$$

$$S = \sqrt{0.273} = 0.5225 \text{ ksi}$$

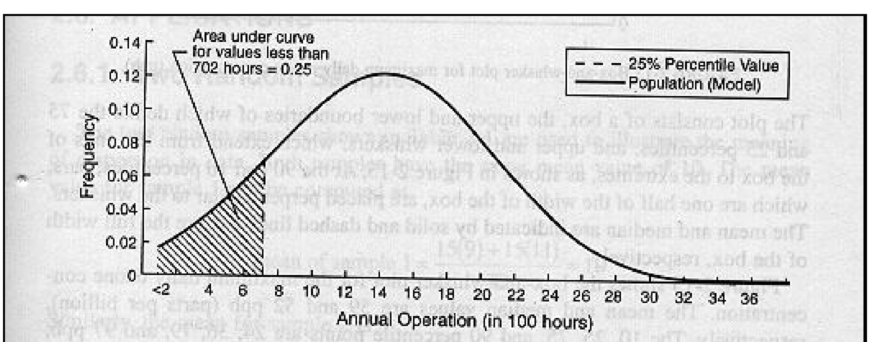
$$\delta \text{ or } COV(X) = \frac{S}{\overline{X}} = \frac{0.5225}{2.86} = 0.183$$

Dispersion Measures

- Percentiles
  - A p percentile value (x_p) for a variable based on a sample is the value of the variable such that p% of the data is less or equal to x_p.
  - On the basis of this definition, the median value is considered to be the 50 percentile value.
  - It is common in engineering to have interest in the 10, 25, 50, 75, and 90 percentile values.

Dispersion Measures

- Example: Operation of a Marine Vessel



Methodology of Modeling Uncertainty

The Methodology of Modeling Uncertainty is described in five chapters that mainly concentrating on how to model uncertainty using probabilities and information as follows:

Detailed Steps

- Probability Basics: reviews fundamental probability concepts. — *Chapter 7*
- Subjective probability: translates beliefs & feelings about uncertainty in probability for use in decision modeling. — *Chapter 8*
- Theoretical Probability Models: helps with representing uncertainty in decision modeling — *Chapter 9*
- Using Data: uses historical data for developing probability distributions — *Chapter 10*
- Monte Carlo Simulation**:** to give the decision-maker a fair idea about the probabilities associated with various outcomes. — *Chapter 11*
- *Value of Information*:explores the value of information within the decision-analysis framework. — *Chapter 12*