IAM 530
ELEMENTS OF PROBABILITY AND STATISTICS

INTRODUCTION

WHAT IS STATISTICS?

- **Statistics** is a science of collecting data, organizing and describing it and drawing conclusions from it. That is, **statistics** is a way to get information from data. It is the science of uncertainty.

WHAT IS STATISTICS?

- Which letter of the alphabet is most frequently used? What are the top 5?

- The chief financial officer in FedEx believes that including a stamped self-addressed (SSA) envelop in the monthly invoice sent to customers will decrease the amount of time it take for customers to pay their monthly bills. Is this really true?

WHAT IS STATISTICS?

– An ergonomic chair can be assembled using two different sets of operations. The operations manager would like to know whether the assembly time under the two methods differ.

WHAT IS STATISTICS?

– The marketing manager needs to decide which of two new packaging designs to adopt, to help improve sales of his company’s soap.
  - Brightly-colored packaging or simple packaging
  - Which packing design must be chosen to increase the sales?

STEPS OF STATISTICAL PRACTICE

- **Model building:** Set clearly defined goals for the investigation
- **Data collection:** Make a plan of what data to collect and how to collect it
- **Data analysis:** Apply appropriate statistical methods to extract information from the data
- **Data interpretation:** Interpret the information and draw conclusions

DESCRIPTIVE STATISTICS AND INFERENTIAL STATISTICS

- **Descriptive statistics** includes the collection, presentation and description of numerical data .
- **Inferential statistics** includes making inference, decisions by the appropriate statistical methods by using the collected data.

BASIC DEFINITIONS

- **POPULATION:** The collection of all items of interest in a particular study.

- **PARAMETER:** A descriptive measure of the population

- **SAMPLE:** A set of data drawn from the population

- **STATISTIC:** A descriptive measure of a sample

- **VARIABLE:** A characteristic of interest about each element of a population or sample.

EXAMPLE 1

Population	Unit	Sample	Variable
All students currently enrolled in school	Student	Any department	GPA Hours of works per week
All books in library	Book	Statistics’ Books	Replacement cost Frequency of check out Repair needs
All campus fast food restaurants	Restaurant	Burger King	Number of employees Seating capacity Hiring/Not hiring

COLLECTING DATA

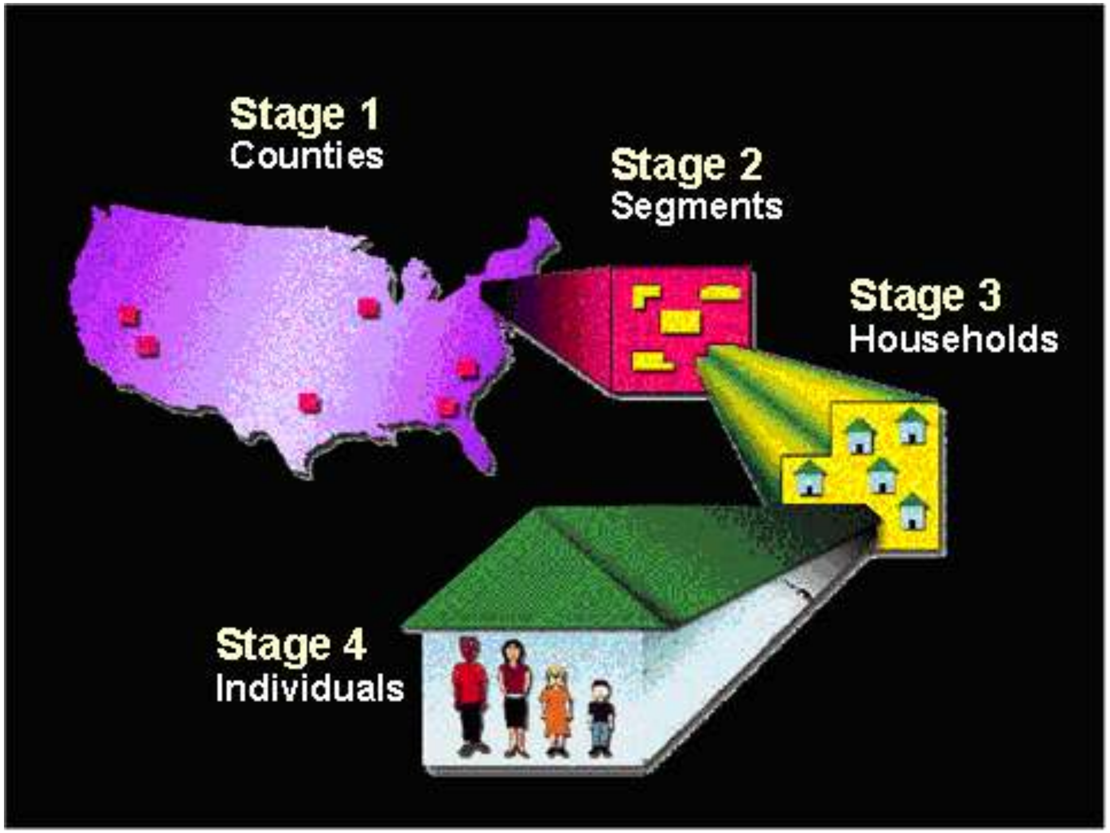
- **Target Population:** The population about which we want to drawn inferences.

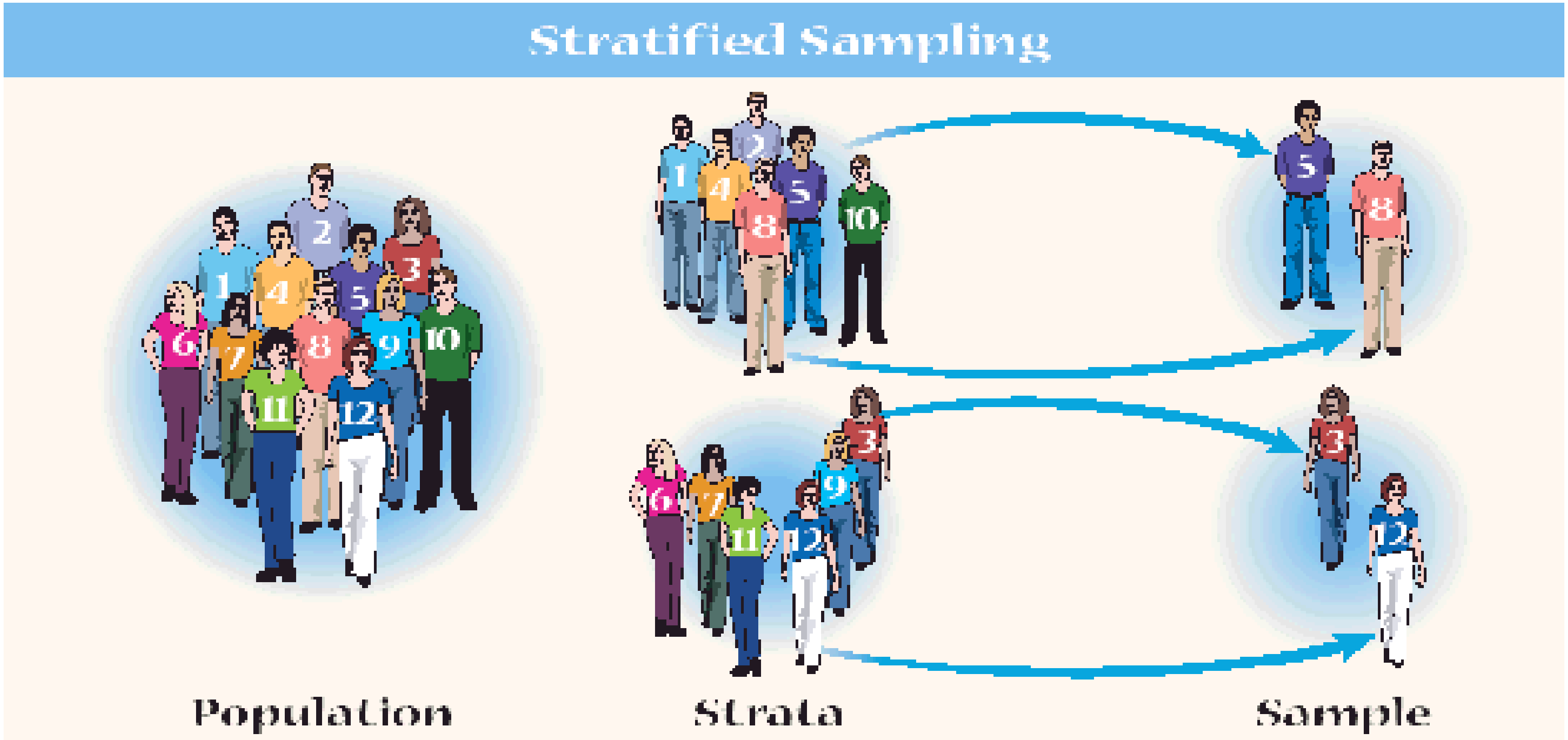
- **Sampled Population:** The actual population from which the sample has been taken.

SAMPLING PLAN

- **Simple Random Sample (SRS):** All possible samples with the same number of observations are equally likely to be selected.
- **Stratified Sampling:** Population is separated into mutually exclusive sets (strata) and then sample is drawn using simple random samples from each strata.
- **Convenience Sample:** It is obtained by selecting individuals or objects without systematic randomization.

Simple Random Sampling (SRS)




Stratified Sampling
Population
Strata
Sample

DESCRIPTIVE STATISTICS

- **Descriptive statistics** involves the arrangement, summary, and presentation of data, to enable meaningful interpretation, and to support decision making.
- **Descriptive statistics** methods make use of
  - graphical techniques
  - numerical descriptive measures.
- The methods presented apply to both
  - the entire population
  - the population sample

Types of data and information

- A variable - a characteristic of population or sample that is of interest for us.
  - Cereal choice
  - Capital expenditure
  - The waiting time for medical services
- Data - the actual values of variables
  - Interval data are numerical observations
  - Nominal data are categorical observations
  - Ordinal data are ordered categorical observations

Types of data - examples

Interval data

Age	income
55	75000
42	68000
.	.
.	

Weight gain
+10
+5
.
.

Nominal

Person	Marital status
1	married
2	single
3	single
.	.

Computer	Brand
1	IBM
2	Dell
3	IBM
.	.
.	.

Types of data - examples

Interval data

Age	**income**
55	75000
42	68000
.	.
.	

Weight gain
+10
+5
.
.

Nominal data

With nominal data, all we can do is, calculate the proportion of data that falls into each category.

IBM	Dell	Compaq	Other	Total
25	11	8	6	50
50%	22%	16%	12%	

Types of data – analysis

- Knowing the **type** of data is necessary to properly select the technique to be used when analyzing data.
- Type of analysis allowed for each type of data
  - Interval data – arithmetic calculations
  - Nominal data – counting the number of observation in each category
  - Ordinal data - computations based on an ordering process

MATHEMATICAL STATISTICS

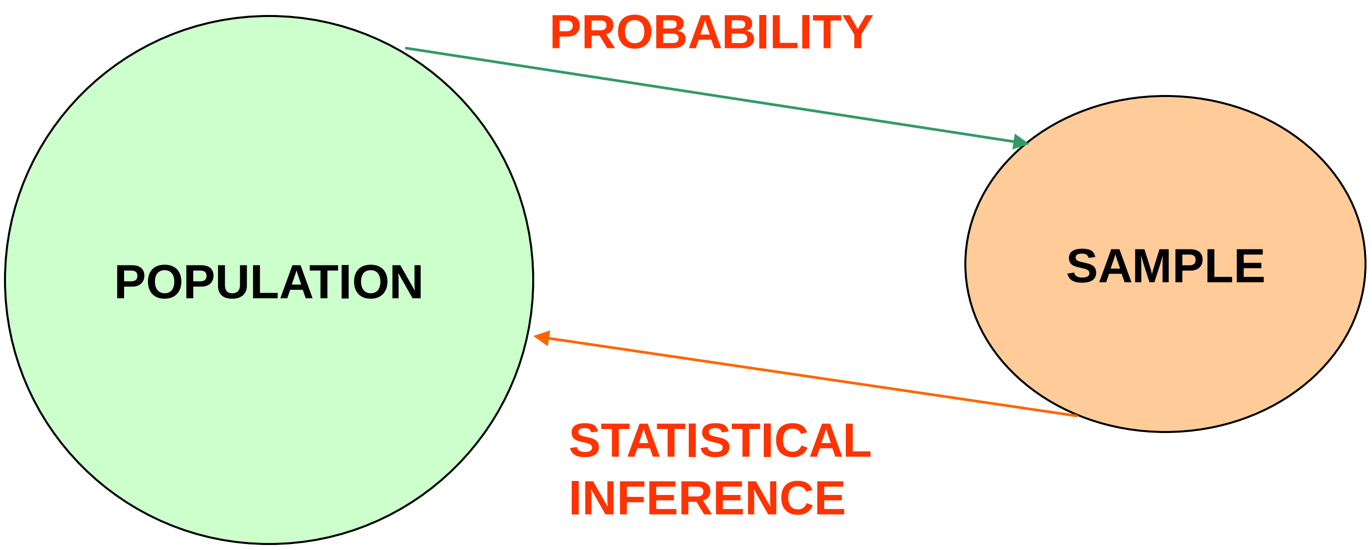
LIFE→UNCERTANTIES →STATISTICS

↨

MATHEMATICS

↓

PROBABILITY THEORY


PROBABILITY
POPULATION
SAMPLE
STATISTICAL
INFERENCE

- **PROBABILITY:** A numerical value expressing the degree of uncertainty regarding the occurrence of an event. A measure of uncertainty.

- **STATISTICAL INFERENCE:** The science of drawing inferences about the population based only on a part of the population, sample.

PROBABILITY

- CLASSICAL INTERPRETATION

  If a random experiment is repeated an infinite number of times, the relative frequency for any given outcome is the probability of this outcome.

  Probability of an event: Relative frequency of the occurrence of the event in the long run.

- SUBJECTIVE INTERPRETATION

  The assignment of probabilities to event of interest is subjective

PROBABILITY

- Random experiment
  - a random experiment is a process or course of action, whose outcome is uncertain.
- Examples

Experiment	*Outcomes*
• Flip a coin	Heads and Tails
• Record a statistics test marks	Numbers between 0 and 100
• Measure the time to assemble a computer	Numbers from zero and above

PROBABILITY

- Performing the same random experiment repeatedly, may result in different outcomes, therefore, the best we can do is consider the ***probability of occurrence of a certain outcome.***

- To determine the probabilities, first we need to define and list the possible outcomes

Sample Space

- Determining the outcomes.
  - Build an exhaustive list of all possible outcomes.
  - Make sure the listed outcomes are mutually exclusive.
- The set of all possible outcomes of an experiment is called a ***sample space*** *and denoted by* Σ*.*

Sample Space

Countable

Finite number of elements

Uncountable

Infinite number of elements

EXAMPLES

- Tossing a coin experiment

$$\Sigma : \{\text{Head, Tail}\}$$

- Rolling a dice experiment

$$\Sigma : \{1, 2, 3, 4, 5, 6\}$$

- Determination of the sex of a newborn child

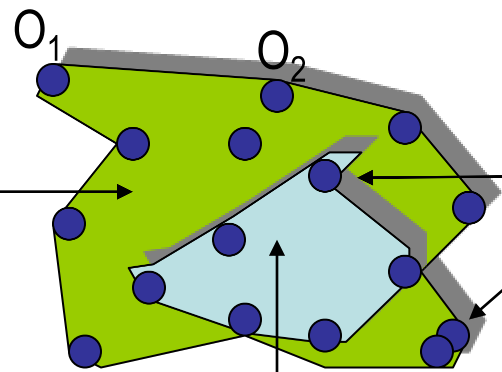
$$\Sigma : \{\text{girl, boy}\}$$

EXAMPLES

- Examine 3 fuses in sequence and note the results of each experiment, then an outcome for the entire experiment is any sequence of N's (non-defectives) and D's (defectives) of length 3. Hence, the sample space is

$$\Sigma : \{ NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

Sample Space: $\Sigma = \{O_1, O_2, \ldots, O_k\}$



Sample Space
a sample space of a random experiment is a list of all possible outcomes of the experiment. The outcomes must be mutually exclusive and exhaustive.

Event
An event is any collection of one or more simple events

Simple Events
The individual outcomes are called *simple events*. Simple events cannot be further decomposed into constituent outcomes.

Our objective is to determine **P(*A*)**, the probability that event ***A*** will occur.

Assigning Probabilities

– Given a sample space $\Sigma = \{O_1, O_2, \ldots, O_k\}$, the following characteristics for the probability $P(O_i)$ of the simple event O_i *must hold:*

$$1. \quad 0 \le P(O_i) \le 1 \quad for \quad each \quad i$$

$$2. \quad \sum_{i=1}^{k} P(O_i) = 1$$

– **Probability of an event**: The probability $P(A)$, of event A is the sum of the probabilities assigned to the simple events contained in A.

Assigning Probabilities

- P(A) is the proportion of times the event A is observed.

$$P(A) = \frac{\text{total outcomes in A}}{\text{total outcomes in S}}$$

•What is the probability of drawing an ace from a well-shuffled deck of 52 playing cards?

Intersection

- The intersection of event A and B is the event that occurs when both A and B occur.

- The intersection of events A and B is denoted by (A and B) or $A \cap B$.
- The joint probability of A and B is the probability of the intersection of A and B, which is denoted by P(A and B) or $P(A \cap B)$.

Union

- The union event of A and B is the event that occurs when either A or B or both occur.

- At least one of the events occur.

- It is denoted “A or B” OR $A \cup B$

Complement Rule

- The complement of event A (denoted by A^C) is the event that occurs when event A does not occur.
- The probability of the complement event is calculated by

A and A^C consist of all the simple events in the sample space. Therefore, $P(A) + P(A^C) = 1$

$$P(A^C) = 1 - P(A)$$

MUTUALLY EXCLUSIVE EVENTS

- Two events A and B are said to be mutually exclusive or disjoint, if A and B have no common outcomes. That is,

$$A \text{ and } B = \varnothing \text{ (empty set)}$$

- The events $A_1, A_2, \ldots$ are **pairwise mutually exclusive** (disjoint), if

$$A_i \cap A_j = \varnothing \text{ for all } i \neq j.$$

MUTUALLY EXCLUSIVE EVENTS

A small city has 3 automobile dealerships: A GM dealer selling Chevrolets, Pontiacs and Buicks; a Ford dealer selling Fords And Mercurys; and a Chrysler dealer selling Plymouths and Chryslers.

If an experiment consists of observing the brand of the next car sold, then events A={Chevrolet, Pontiac, Buick} and B={Ford, Mercury}

are **mutually exclusive.** Why?

EXAMPLE 2

- The number of spots turning up when a six-sided die is tossed is observed. Consider the following events.

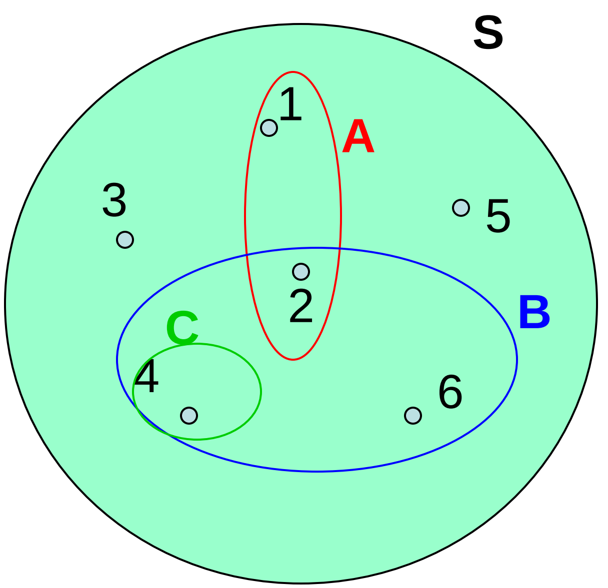
A: The number of observed is at most 2.

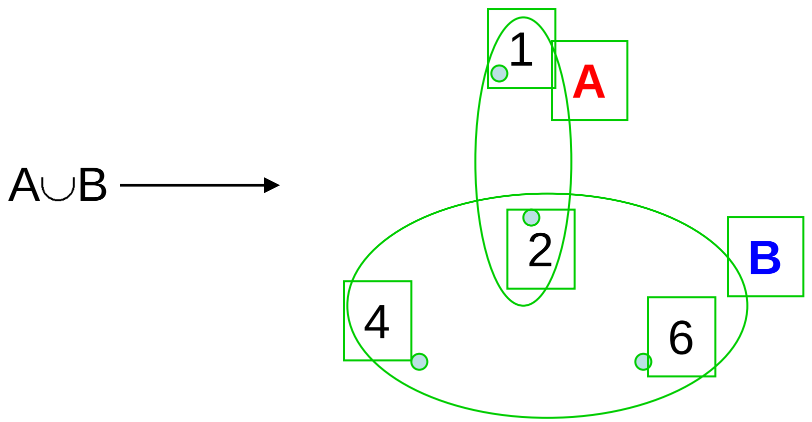
B: The number observed is an even number.

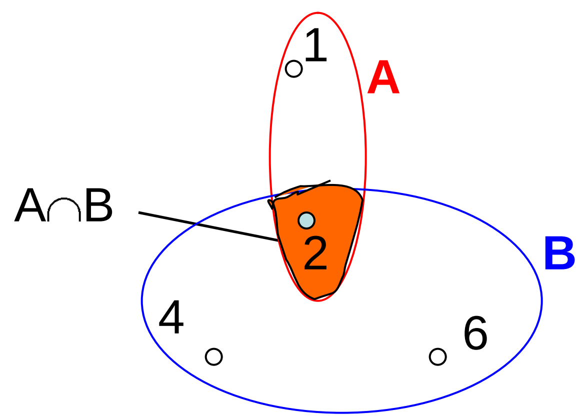
C: The number 4 turns up.

VENN DIAGRAM

- A graphical representation of the sample space.







$A \cap C = \varnothing \rightarrow$ A and C are mutually exclusive

AXIOMATIC FOUNDATIONS

Borel Field (Sigma-Algebra) denoted by $\mathcal{B}$:

A collection of subsets of the sample space Σ satisfying

a) $\emptyset \in \mathcal{B}$

b) If $A \in \mathcal{B}$, then $A^C \in \mathcal{B}$

c) If $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

Borel Field (Sigma-Algebra)

- If Σ is finite or countable, then

  B = {all subsets of Σ, including Σ itself}

  Example: $\Sigma = \{1,2,3\} \rightarrow \mathrm{B}$ is the collection of $2^3=8$ sets

$\{1\}$	$\{1,2\}$	$\{1,2,3\}$
$\{2\}$	$\{1,3\}$	$\varnothing$
$\{3\}$	$\{2,3\}$	

AXIOMS OF PROBABILTY (KOLMOGOROV AXIOMS)

Given a sample space Σ and an associated Borel field B, the **probability function** is a function P with domain B that satisfies

1) For any event A, $0 \le P(A) \le 1$.
2) $P(\Sigma) = 1$.
3) If $A_1, A_2, \ldots \in \mathrm{B}$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{\infty} P(A_i),\ n = 1, 2, \ldots$$

THE CALCULUS OF PROBABILITIES

- If P is a probability function and A is any set in $\mathcal{B}$, then

a. $P(\emptyset)=0$

b. $P(A) \leq 1$

c. $P(A^C)=1 - P(A)$

THE CALCULUS OF PROBABILITIES

- If P is a probability function and A and B any sets in $\mathcal{B}$, then

a. $P(B \cap A^C) = P(B) - P(A \cap B)$

b. If $A \subset B$, then $P(A) \leq P(B)$

c. $P(A \cap B) \geq P(A) + P(B) - 1$ (Bonferroni Inequality)

d. $\boldsymbol{P}\left(\bigcup_{i=1}^{\infty} \boldsymbol{A}_i\right) \leq \sum_{i=1}^{\infty} \boldsymbol{P}\left(\boldsymbol{A}_i\right)$ for any sets $A_1, \boldsymbol{A}_2, \cdots$

(Boole's Inequality)

EXAMPLE 3

- A student has a box containing 25 computer disks of which 15 are blank and the other 10 are not. If she randomly selects disks one by one, what is the probability that at least 2 must be selected in order to find one that is blank?

EQUALLY LIKELY OUTCOMES

- The same probability is assigned to each simple event in the sample space, Σ.
- Suppose that $\Sigma=\{s_1,\ldots,s_N\}$ is a finite sample space. If all the outcomes are equally likely, then $P(\{s_i\})=1/N$ for every outcome s_i.

Then, for any event A,

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S}$$

Addition Rule

For any two events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Suppose that a metal fabrication process yields output with faulty boundings (FB) 10% of the time and excessive oxidation (EO) in 25% of all segments, 5% of the output has both faults. What is the probability that either of the events will occur?

ODD EVENTS

- The odd event of A is defined by

$$\frac{P(A)}{P(A^C)} = \frac{P(A)}{1 - P(A)}$$

•It tells us how much more likely to see the occurrence of event A.

•P(A)=3/4→P(A^C)=1/4→ P(A)/P(A^C) = 3. That is, the odds is 3. It is 3 times more likely that A occurs as it is that it does not.

COUNTING TECHNIQUES

- Methods to determine how many subsets can be obtained from a set of objects are called counting techniques.

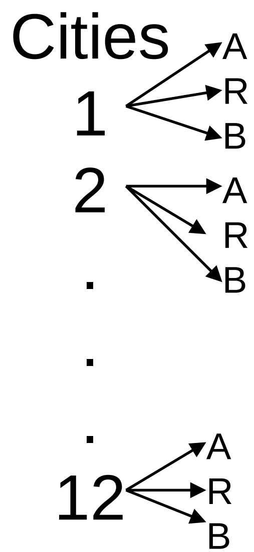
FUNDAMENTAL THEOREM OF COUNTING

If a job consists of k separate tasks, the i-th of which can be done in n_i *ways*, $i=1,2,\ldots,k$, then the entire job can be done in $n_1 x n_2 x \ldots x n_k$ ways.

EXAMPLE 4

- A travel agency offers weekend trips to 12 different cities by air, rail or bus. In how many different ways such a trip be arranged?

Tree Diagram:



EXAMPLE 5

- If a test consists of 12 true-false questions, in how many different ways can a student mark the test paper with one answer to each question?

THE FACTORIAL

- The number of different sequences that might be possible when observations are taken one at a time.

$$n! = n(n-1)(n-2)\ldots 2.1$$

$$0! = 1, \; 1! = 1$$

Example: Among 4 people how many different ways are there to choose a president, a vice president, a treasurer and a secretary?

COUNTING

- In the New York State lottery, the first number can be chosen 44 ways. There are 6 numbers on the tickets.

1. Ordered, without replacement
2. Ordered, with replacement
3. Unordered, without replacement
4. Unordered, with replacement

COUNTING

- **Partition Rule:** There exists a single set of N distinctly different elements which is partitioned into k sets; the first set containing n_1 elements, …, the *k*-the set containing n_k elements. The number of different partitions is

$$\frac{N!}{n_1!n_2!\cdots n_k!} \text{ where } N = n_1 + n_2 + \cdots + n_k.$$

COUNTING

	Number of possible arrangements of size r from n objects	
	Without Replacement	With Replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

EXAMPLE 6

- A class in statistics consists of 6 men and 4 women. An examination is given and the students are ranked according to their performances. Assume that no students obtain the same score,

a) How many different rankings are possible?

b) If the men are ranked among themselves and the women are ranked among themselves, how many different rankings are possible?

EXAMPLE 7

- How many ways can the letter in the word COMPUTER be arranged in a row?
- In how many arrangements O is just after C?
- How many start with CO?

PERMUTATIONS

- Any ordered sequence of r objects taken from a set of n distinct objects is called a **permutation** of size r of the objects.

$$P_{r,n} = \frac{n!}{(n-r)!} = n(n-1)...(n-r+1)$$

- Both composition and order are important.

EXAMPLE 8

- Suppose that a random sample of 5 students is taken one at a time without replacement from the 38 members of the Tau Beta Pi. The number of sample outcomes is

EXAMPLE 9

- Consider the case where 5 of 7 students are to be seated in a row. In how many ways these students can be seated?

COMBINATION

- Given a set of n distinct objects, any <u>unordered</u> subset of size r of the objects is called a **combination**.

$$C_{r,n} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Properties

$$\binom{n}{0} = 1, \qquad \binom{n}{n} = 1, \qquad \binom{n}{r} = \binom{n}{n-r}$$

EXAMPLE 10

- In how many different ways can 3 of 20 laboratory assistants be chosen to assist with an experiment?

EXAMPLE 11

- A carton of 12 rechargeable batteries contains one that is defective. In how many ways can an inspector choose 3 of the batteries and

a) Get the one that is defective?

b) Not get the one that is defective?

EXAMPLE 12

- We have 24 light bulbs, 2 defective. If 2 of the bulbs are chosen at random what are the probabilities that

a) Neither bulb will be defective?

b) One of the bulb will be defective?

c) Both bulbs will be defective?