INTRODUCTION

"Statistical thinking will in one day be as necessary as the ability to read and write."
- *by H.G. Wells*

We have come into an age of computerization and henceforth we are becoming rich in information. However, these data will not make sense unless we know how to use the available information to make better decisions. This problem can be aided by statistics because statistics deals with the collection, presentation, analysis and interpretation of a set of data in order to yield meaningful information.

Uses of Statistics:

1. to know how to properly present and describe information
2. to know how to draw conclusion about large populations based only on information obtained from samples.
   Population – refers to the totality of the observations of which the study is concerned
   Sample – refers to a part or subset of a population
3. to know how to improve processes like strategies how to improve sales or quality of a product or improve services delivered by an organization
4. to know how to obtain reliable forecasts

Two Major Areas of Statistics:

1. **Descriptive statistics –** a branch of statistics concerned with numerical and graphical techniques fro **describing** one or more characteristics of a group and comparing characteristics between groups.
2. **Inferential statistics –** a branch of statistics concerned with the acquisition of data, with sampling, and with the use of the data in making **inferences (conclusion)** about a population.

Parameter – is a summary measure that describe a characteristic of an entire population and this measure is computed using the entire population

Statistic – is a summary measure that is computed based from a sample of the population and is used to describe a characteristic of the sample

Example. Suppose a study will be conducted by the president of LSU in order to learn about student perceptions concerning the imposition of a tuition fee increase.

Population of the study: All the currently enrolled students of LSU.

Main objective of the study: To estimate various attitudes or characteristics of interest of the entire population.

Application of Inferential statistics: Select only a sample of the population and use the statistics computed from the sample to draw conclusions about the population parameters or characteristics.

Remark: Inferential statistics have been developed due to the need of studying only a sample instead of the whole population.

Advantages of studying a sample:

Because lesser respondents or experimental units will be involved then sampling is better because

1. it entails lesser cost
2. it is less time-consuming
3. it is less cumbersome and more practical to administer
4. some experiments are destructive so it is not possible to involve the whole population. For example in the quality sampling of products, if you test or taste the product, usually you cannot sell it anymore.

Type of Observations

a) **Constant** – occurs when the phenomenon or value of a sample or population will **not change** after successive trials. Examples: tribe of a Maranao family,
b) **Variable** – occurs when the phenomenon or value of a sample or population will **change** after successive trials. Examples: tribes of people living in Marawi City, family income, age, birth rate.

Types of Variable

a) **Quantitative variable** – one whose observation vary in magnitude from trial to trial. (greatness of size, degree, quantity). Examples: family income, population size, age.
   i.) **Discrete quantitative variable** – a quantitative variable that can assume only a countable number of values. Examples: number of children for each household, number of deaths for each year due to lung cancer.
   ii.) **Continuous quantitative variable** - a quantitative variable can assume any one of the countless number of values in a line interval. Example: height of a person, weight of a lady, temperature of a certain room, time.
b) **Qualitative variable** - one whose observation vary in kind (a group united by common traits or interest) but not in degree. Example: Sex, Religion, Marital Status, Political Party.

Scales(Levels) of Measurement (Data)

a) **Nominal Scale** – defines specific categories by name. These categories are called **levels** of the scale. **All qualitative variables are measured on a nominal scale**. Example: Political Party, Sex and others.
b) **Ordinal Scale** – incorporates the features of a nominal scale and additional feature that observations can be **ordered or ranked** from low to high. Example: categorize gross annual income, prestige ranking of occupations.
c) **Interval Scale** – incorporates all features of an ordinal (and hence nominal)scale and the additional feature that we can specify **distances** between levels on the scale. Example: IQ
d) **Ratio Scale** – incorporates all the features of interval (and hence nominal and ordinal) scales and the additional feature that ratios can be formed with levels of the scale, and there is a meaningful zero point for the scale. Example: Temperature, birth rate

Type of numerical descriptive measures

a.) **Measures of Central Tendency**- measures that can possibly locate the center of the distribution data..

i.) **Mode** – the measurement that occurs with greatest frequency (**occurs frequently**). Sometimes a set of measurements has more than one mode (Bimodal, Trimodal and so on).

ii.) **Median** – a number chosen so that half the measurements lie below it, half above.

- For **odd number of measurements** the median is the middle measurement when the measurement are arranged in order of magnitude. $median = \frac{n+1}{2}$ measurement.

Example: $62, 73, 78, 86, 89, 90, 95$ (7 observations therefore odd)

Median is the $\frac{7+1}{2}$ th = 4th observation which is 86.

- For **even number of measurement** the median is the average of the two middle observation when the measurements are arranged in order of magnitude.

Example: $62, 73, 73, 75, 78, 86, 89, 90, 91, 95$ (10 observations therefore even)

The two midpoint scores are 78 and 86; hence, the median is $median = \frac{78+86}{2} = 82$

iii.) **Mean** (arithmetic mean) – is the sum of a set of measurements divided by the number of measurements in the set. The average of the set of measurements.

Example: $95, 86, 78, 90, 62, 73, 89$

Mean = $\frac{sum\ of\ scores}{n} = \frac{95+86+78+90+62+73+89}{7} = 81.86$

$\overline{X}$ **is the sample mean**

μ **is the population mean**

$$\overline{X} = \frac{\sum X}{n} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

b) Measures of Variation (Dispersion) – measures that describe the spread of the distribution data.

i.) **Range** – defined to be the difference between the largest and the smallest measurement.

Example: $95, 86, 78, 90, 62, 73, 89$ Range = $95 - 62 = 33$.

ii.) **Variance** – defined to be the sum of the squared deviations of the measurements about their mean, divided by n-1 for a sample.

$\frac{\sum\left(X - \overline{X}\right)^2}{n-1} = s^2$ **-the sample variance**

$\frac{\sum\left(X - \overline{X}\right)^2}{n} = \sigma^2$ **-the population variance**

Example: $68, 67, 66, 63, 61$

Measurement X	Deviation $X-\overline{X}$	Square of the deviation $\left(X-\overline{X}\right)^2$
68	$68-65=3$	9
67	$67-65=2$	4
66	$66-65=1$	1
63	$63-65=-2$	4
61	$61-65=-4$	16

$$\overline{X}=\frac{68+67+66+63+61}{5}=65$$

$$s^2=\frac{\sum\left(X-\overline{X}\right)^2}{n-1}$$

$$s^2=\frac{9+4+1+4+16}{5-1}=\frac{34}{4}=8.5$$

Shortcut formula: $s^2=\frac{1}{n-1}\left[\sum X^2-\frac{\left(\sum X\right)^2}{n}\right]$

Measurement X	Square of X X^2
68	4624
67	4489
66	4356
63	3969
61	3721
$\sum X=325$	$\sum X^2=21159$

$$s^2=\frac{1}{5-1}\left[21159-\frac{(325)^2}{5}\right]=\frac{1}{4}\left[21159-\frac{105625}{5}\right]=\frac{1}{4}\left[21159-21125\right]=\frac{34}{4}=8.5$$

iii.) **Standard Deviation** – Defined to be the positive square root of the variance.

s is the sample standard deviation:

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

σ is the population standard deviation:

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

iv.) **Coefficient of relative variation (CRV)** – standard deviation expressed as a percentage of the mean. $CRV=\left(\frac{s}{\overline{X}}\right)\times 100$

However, the biggest disadvantage of sampling is that the sample may not truly reflect the characteristic of the population which would lead to wrong conclusions. Hence, care must be taken in choosing your sample.

Types of samples:

1. **Probability sample** – is one in which the elements of the sample are chosen o the basis of known probabilities. Each element in the population have an equal and independent chance of selection in the sample. This means that the choice of an element is not influenced by other considerations such as personal preference, and that the choice of one element is not dependent upon the choice of another element in the sampling.

2. **Non-probability sample** – is one in which individuals or items are chosen without regard to their probability of occurrence. This is usually used when the number of elements of the population is either unknown or cannot be individually identified.

SAMPLING PROCEDURES

I. **Non-Probability Sampling**

Some members in the populations are given zero chance of being included in the sample.

1) *Purposive Sampling*- making a sample which agrees with the profile of the population based on some pre-selected characteristics.

2) *Quota Sampling*-selecting a specified number (quota) of units possessing certain characteristics.

3) *Convenience Sampling*- using results that are readily available.

4) *Judgement Sampling*- selecting a sample in accordance with an expert's judgement.

II. **Probability Sampling**

Every member in the population is given equal chance of being included in the sample.

1) *SRS (Simple Random Sampling)*

Done with or with out replacement.

Procedure:

Step #1: Number the elements of the population from 1 to N.

Step #2: Select n numbers from 1 to N using a random process.

2) *Systematic Random Sampling*

Select every kth element in the population, the first unit being selected at random.

Procedure:

Step #1: Number the population units from 1 to N.

Step #2: Determine the Sampling interval, k.

$k = N/n$ where N- population size, n- sample size

Step #3: Select a random start, r. $1 \le r \le k$

(Note: the first unit of the sample is the unit corresponding to r.

3) *Stratified Random Sampling*

The population of N units is divided into homogeneous subpopulations (called strata) and then a sample is drawn from each stratum.

Procedure:

Step #1: Classify the population into at least two homogeneous strata.

Step #2: Using proportional allocation, draw a sample from each stratum.

Proportional allocation- the process of determining the number of units to be taken from each stratum.

If the size of the population N is divided into k homogeneous subpopulations or strata of sizes N_1, N_2, …, N_k, then the sample size to be taken from each stratum i is obtained using the formula:

$N = (n/N) N_i$, $i = 1, 2, \ldots, k$

4) *Cluster Sampling*

This method selects a sample containing either all, or a random selection, of the elements from clusters that have themselves been selected randomly from the population.

Procedure:

Step #1: Divide the population area into sections or clusters (heterogeneous)

Step #2: Select randomly a few from these clusters.

Statistical Experiment and Counting Techniques

Statistical Experiment – is an experimental that generates data. Example: tossing of coin, tossing a die, drawing a ball from a box.

Sample Space – the set of all possible outcomes of an experiment, denoted be S.

Sample Point – refers to each outcome in the sample space.

Events – a subset of the sample trace, denoted by capital letters A, B, C,…., Z.

1. Simple event – a set containing only one element in the sample space.
2. Compound event – one that can be expressed as a union of simple events.

Null or empty space – a set that contains no element, denoted by Ø.

Disjoint sets/ mutually exclusive events – sets which do not have elements in common.

Proportion – the number of elements in a set possessing a specific characteristic divided by the total number of elements in the set.

Percentage – a proportion multiplied by 100.

Rate of occurrence – The number of actual occurrences divided by the number of possible occurrences.

In many cases, we shall be able to count the number of points in the sample space without actually listing element.

Three Counting Principles:

A. FPC (Fundamental Principle of Counting)

If an operation can be perform can be performed in n_1 ways and of for each of these a second operation can be done in n_2 ways, then the two operations can be done in $n_1 \times n_2$ ways.

Examples: 1. How many sample points are there in the sample space when a pair of dice is thrown once?

Solution:

The first die can land in any of 6 ways. For each of these 6 ways the second die can also land in 6 ways. Therefore, the pair of dice can land (6)(6) = 36 ways.

2. How many sample points are in the sample space when a coin is tossed three times?

Solution:

(2) (2) (2) = 8 sample points

B. **Permutation** – An ordered arrangement of all or part of n distinct objects.

Property 1: The number of permutations of n distinct objects taken all at a time is n!

Examples:

1) How many distinct permutations can be made from the lettersof the word LOVE?

Solution:

n! = 4! = 4 x 3 x 2 x 1 = 24 ways (there are 4 letters that can be chosen for the first position, the next position can be used by any of the three remaining letters, and so on, until the last position is filled by the remaining letter. Some of these are LOVE, LOVE, LVEO, etc.

2) In how many ways can An, Joe, Al, Fe, and Cale be seated in a row of five chairs?

Solution:

n! = 5! = 5 x 4 x 3 x 2 x 1 = 120 ways.

Hence, from the fundamental principle of counting the number of committees that can be formed with 2 mathematicians and 1 statistician is n x n = 6×3 = 18 ways.

Property 2: The number of permutations of n distinct objects taken r at a time is

$$ {}_nP_r = \frac{n!}{(n-r)!} $$

Examples:

1) Suppose there are 4 objects {a, b, c, d}. If we arrange these objects two at a time, how many possible arrangements are there?

Solution:

n = 4, r = 2, $\frac{4!}{(4-2)!} = \frac{4!}{2!} = \frac{4 \times 3 \times 2!}{2!} = 12$ possible permutations or arrangements. These are ab, ac, ad, ba, ca, ba, bc, bd, cd, cb, db, & dc.

2) If two prizes, the first and second price, will be awarded among 10 students, how many ways can this be done assuming that only one prize will be won by a winner?

Answer: 90 ways

Property 3: The number of permutations of n distinct object arranged in a circle is $(n-r)!$

Example: In how many ways can 6 different varities of orhids be planted in a circle?

Solution:

There are n = 6 varieties of orchid plants. The number of permutations is (6-1)! = 5! = 5 x 4 x 3 x 2 x 1 = 120 ways that these plants will be planted in a circle.

Property 4: The number of distinct permutations of n things of which n_1 ***are of one kind,*** n_2 ***of a second kind (or alike),...*** n_k ***of the kth kind, is given by*** $\frac{n!}{n_1!n_2!\ldots n_k!}$ ***where*** $n = n_1 + n_2 + \ldots + n_k$

Example: How many different ways can 3 red, 4 yellow, and 2 green bulbs be arranged in a string of Christmas lights with 9 sockets?

Solution:

The total number of distinct arrangements is

$$\frac{9!}{3!4!2!} = \frac{9\times8\times7\times6\times5\times4!}{3!4!2!} = 1{,}260 \text{ ways.}$$

C. COMBINATION – the number of ways of selecting r objects without regard to order. Property: The number of combinations of n distinct objects taken r at a time is ${}_nC_r = \frac{n!}{(n-r)!r!}$

Examples:

1. How many different combinations can be formed from the letter a, b, c, and d if two letters are taken at a time?

   Solution:

   n = 4, r = 2, the total number of combinations is ${}_4C_2 = \frac{4!}{(4-2)!2!} = 6$ ways.

2. From 4 mathematicians and 3 statisticians, find the number of committee of size three that can be formed with two mathematicians and one statistician.

   Solution:

   The number of ways n_1, to select two mathematicians from 4 is ${}_4C_2 = \frac{4!}{(4-2)!2!} = \frac{4\times3\times2!}{2!2!} = 6$ ways

   The number of ways n_2, to select one statistician from 3 statisticians is

   $${}_3C_1 = \frac{3!}{(3-1)!1!} = \frac{3\times2!}{2!} = 3 \text{ ways}$$

Using the fundamental principle of counting, the number of committee of size three that can be formed with two mathematicians and one statistician is

$n_1 \times n_2 = 6\times3 = 18$ ways.

Exercises:

1. I f an experiment consists of throwing a die and then drawing a letter at random from the English alphabet, how many points are in the sample space?
2. A college student must take a science course, an English course, and a statistics courses, and any of 3 statistics courses, how many ways can she arrange her program?
3. How many distinct permutations can be made from the letters of the word "columns"? How many of there permutations starts with letter "n"?
4. How many ways can 6 people be lined up in a bus?
5. How many ways can a caravan of 8 covered wagons from Arizona be arranged in a circle?
6. From a group of 4 men and 5 women, how many committees of size 3 are possible a) with no restriction?
   b) with 1 man and 2 women?
   c) with 3 men and no women?
7. How many ways can a local chapter of the mathematical society of the Philippines schedule three speakers for three different seminars if they are all available on any of five possible dates?
8. Find the number of ways in which six teacher can be assigned to four sections of an introductory psychology course if no teacher is assigned to more than one section.

PROBABILITY

In the 17^{th} century, two mathematicians, Blaise Pascal and Pierre de Fermat carried on a correspondence discussing mathematical problems dealing with games of chance. They were confronted with questions raised by gamblers: How can the uncertainty of the outcomes of a game, such as coins, dice or deck of cards, be measured? What constitute a fair bet? Attempts to answer these questions led to the birth of probability theory.

Probability is a branch of mathematics that deals with calculating the occurrence of a given event.

Properties of Probability

1. $0 \le P(A) \le 1$
2. If A= Ø, then P(A) = 0. (impossibility)
3. If A = S, then P(S) = 1. (certainty)

Approaches of Probability

1. **Subjective**

- the probability depends on one's personal assessment of how likely an event will occur.
- the use of personal beliefs, intuition and other indirect information in arriving probabilities are under subjective approach.

Example: A Lakers supporter might say, "I believe that Lakers have probability of 0.9 of winning the NBA this year since they have been playing really well."

2. **Empirical**

- the probability of an event A can be approximated by the proportion of times that A occurs when the experiment is repeated a very large number of times.
- it is expressed in terms of relative frequencies.

Examples

1. If a coin is tossed then, what is the probability that the coin will turn up head? Suppose that the coin is tossed 100 times under the same conditions. Then suppose that the coin fall heads 45 times out of 100.

$$P(A) = \frac{45}{100} = \frac{9}{20}.$$

2. Consider Table 1.3. Find the probability that the observation belongs to the third interval.

Frequency Distribution Table of Weights of Math 31 Students

CI	CB	f	CM
40 - 46	39.5 – 46.5	6	43
47 – 53	46.5 – 53.5	14	50
54 - 60	53.5 – 60.5	10	57
61 - 67	60.5 – 67.5	6	64
68 – 74	67.5 – 74.5	2	71
75 - 81	74.5 – 81.5	2	78

$$P(A) = \frac{10}{40} = \frac{1}{4}.$$

3. **Classical**

- probability is computed based on theoretical assumptions about the possible outcomes. Assume that all the possible outcome of an experiment is equally likely. That is, it has the same chance of occurring.

The probability if an event A is equal to the number of possible outcomes, favorable to A divided by the total number of outcomes of the experiment. That is,

$$P(A) = \frac{n(A)}{n(S)}.$$

Examples

1. If a die is tossed, each of the 6 numbers should be considered equally likely to occur. Then, the probability that any number (say 4) will occur is 1/6.
2. Drawing a vowel in the English Alphabet.
   P(A) = 5/26

Some Laws of Probability

1. If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

Mutually Exclusive Events - the events do not have sample points in common, $A \cap B = \varnothing$. That is, the events do not occur simultaneously.

The probability of occurrence of either A or B is the sum of their separate probabilities.

$\cup$ in Layman's term is "at least", it is also applicable in solving probabilities involving "either or".

Examples

1. In tossing a die. Let E be the event that even numbers will occur and B be the event that odd numbers will occur.

S = {1, 2, 3, 4, 5, 6}
A = {2, 4, 6}
B = {1, 3, 5}
$P(A \cup B) = P(A) + P(B)$ = 3/6 + 3/6 = 1

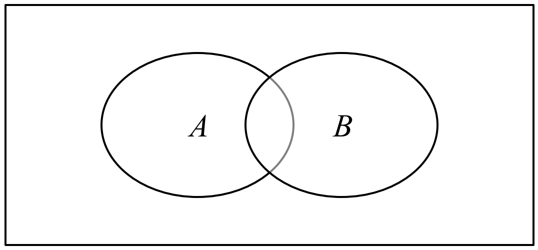
2. If a card is drawn from a deck of cards, what is the probability that it will be either an ace or a queen?

P(ace) = 4/52

P(queen) = 4/52

P(ace or queen) = 4/52 + 4/52 = 8/52 = 2/13.

3. If A and B are any events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



In the Venn diagram above, $P(A \cup B)$ is the sum of the probabilities of the sample points $A \cup B$. Now, P(A) + P(B) is the sum of all probabilities of the points in A and the sum of all probabilities in B. The overlapping events, P(A) + P(B) includes the probabilities of $A \cap B$ twice. Thus, if A and B are joint events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

$\cap$ in Layman's term is "both" or "common" to the given events

Example

1. If a card is drawn from a deck of cards, what is the probability that it will be either heart or a king?

P(heart) = 13/52

P(king) = 4/52

P(heart and king) = 1/52

P(heart or king) = 13/52 + 4/52 – 1/52 = 16/52 = 4/13.

4. If A is an event, then P(A') = 1 - P(A).

The complement of an event A, denoted by A', is the set of all points in S but not in A.

Example

1. What is the probability of obtaining 1, 2, 3, 4 or 5, when an ordinary die is tossed?

P(1, 2, 3, 4 or 5) = 1 – P(6) = 1 – 1/6 = 5/6.

2. Tossing 2 dice, what is the probability that the sum of points will not be 5?

P(5') = 1 – P(5) = 1 – 4/36 = 32/ 36

Exercises

1. a. Suppose a coin is tossed three times. What is the probability of getting 2 heads?

b. Suppose the experiment is performed 100 times and the frequency for each outcome that turned up are as follows:

Outcome	*HHH*	*HHT*	*HTH*	*HTT*	*THH*	*THT*	*TTH*	*TTT*
Frequency	11	5	13	9	17	15	18	12

What is the probability of getting 2 heads?

2. If A and B are mutually exclusive events and P(A) = 0.3 and P(B) = 0.5, find

a. $P(A \cup B)$ b. P(A') c. $P(A' \cap B)$

Hint: Construct a Venn diagram and fill in the probabilities associated with the regions.

3. A pair of dice is tossed. Find the probability of getting a total of 8.

4. In a college graduating class of 100 students, 54 studied mathematics, 69 studied history and 35 studied both mathematics and history. If one of these students is selected at random, find the probability that a. the student takes mathematics or history and b. the student does not take any of these subjects

5. A survey of a class of 35 students in a business school showed the following selection of majors:

Accounting	10
Finance	5
Info Tech	3
Management	6
Marketing	10

Suppose you select a student and observe his or her major, what is the probability that he or she is a Management major?

Conditional Probability

Definition: Let A1 and A2 be events such that P(A1) ≠ 0. The conditional probability of A2 given that A1 has occurred, denoted by P(A2/A1), is defined by

$$P(A2/A1) = \frac{P(A1 \cap A2)}{P(A1)}$$ =probability of both events / prob of given event

Ex. It is estimated that 15% of the adult population has hypertension, but that 75% of all adults feel that personally they do not have this problem. It is also estimated that 6% of the population has hypertension but does not think that the disease is present.

a.) If an adult patient reports thinking that he or she does not have hypertension, what is the probability that the disease is, in fact, present?

b.) If the disease is present, what is the probability that the patient will suspect its presence?

Definition: Two events are independent if either P(A/B) = P(A) or P(B/A) = P(B). Otherwise, A and B are dependent.

Ex. Suppose two cards are drawn in succession from an ordinary deck with replacement.

a.) Let A be the event that the first card is an ace and B be the event that the second card is a spade. Are A and B independent events? (mutually exclusive events)

b.) Let A be the event that the first card is a king and B the event that the second card is a spade. Are A and B independent? (not mutually exclusive events)

Exercises:

1. Suppose a family has four children.
    a.) Find the probability that exactly two are male.
    b.) What is the probability that exactly two are male if the first child born is male?
    c.) Find the probability that the last child born is a male?

d.) What is the probability that the last child born is male if the first three are female?

2. In a study of alcoholics, it was found that 40% had alcoholic fathers and 6% had alcoholic mothers. Forty-two percent had at least one alcoholic parent. What is the probability that a randomly selected alcoholic will
   a.) have both parents alcoholic?
   b.) Have an alcoholic mother if the father is alcoholic?
   c.) Have an alcoholic mother but not an alcoholic father?
   d.) Have an alcoholic mother if the father is not alcoholic?

Multiplicative Rules

Theorem. (Multiplicative Rule) If in an experiment the events A and B can both occur then

$$P(A\cap B) = P(A)\ P(B/A) \quad \text{or} \quad P(A\cap B) = P(B)\ P(B/A).$$

And if events A and B are independent, then $P(A\cap B) = P(A)\ P(B)$.

Ex. In the previous example on hypertension, how many percent of the population has hypertension and they know about its presence?

Ex. If we choose three women from the population and each one becomes pregnant, what is the probability that all three children will be girls?

Ex. Suppose we have a fuse box containing 20 fuses, of which five are defective. If 2 fuses are selected at random and removed from the box in succession without replacing the first, what is the probability that both fuses are defective?

Theorem. (Generalized Multiplicative Rule) If in an experiment the events $A_1, A_2, \ldots, A_k$ can occur, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)\cdots P(A_k \mid A_1 \cap A_2 \cap \cdots \cap A_{k-1})$$

If the events $A_1, A_2, \ldots, A_k$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2)P(A_3)\cdots P(A_k)$$

Ex. The probability that a person visiting his dentist will have an X-ray is 0.6; the probability that a person who has an X-ray will also have a cavity filled is 0.3; and the probability that a person who has had an X-ray and a cavity filled will also have a tooth extracted is 0.1. What is the probability that a person visiting his dentist will have an X-ray, a cavity filled and a tooth extracted?

Ex. Three cards are drawn in succession from an ordinary deck of playing cards. Find the probability that the first card is a red ace, the second card is a ten or jack, and the third card is greater than 3 but less than 7 if the cards are drawn

a.) without replacement.

b.) with replacement.

Exercise:

1. The probability that a doctor correctly diagnoses a particular illness is 0.7. Given that a doctor makes an incorrect diagnosis, the probability that the patient enters a lawsuit is 0.9. What is the probability that the doctor makes an incorrect diagnosis and the patient sues?
2. A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the ambulance and the fire engine will be available.

Diagnostic Tests

Diagnostic Test – a test given to detect the presence of some specific condition in an experimental unit

Four possible conditions in a diagnostic test:

1. The condition is present and the test detects its presence – true positive result, no error is made.
2. The condition is present but the test does not detect its presence – a false negative result, error is made.
3. The condition is not present but the test detects its presence – a false positive result, error is made.
4. The condition is not present and the test does not indicate its presence – a true negative result, no error is made.

Two error rates – probabilities of committing error in a diagnostic test

1. False-positive rate – denoted by α is given by

   α = P(test results are positive / subject is a true negative)

2. False-negative rate – denoted by β is given by

   Β = P(test result is negative / subject is a true positive)

Ex. The serum of a pregnant woman can be analyzed by using a procedure known as starch gel electrophoresis. This test may reveal the presence of a protein zone called the pregnancy zone which is thought to be an indicator that the child is female. To investigate the properties of this test 300 women were selected for study. The result of the test and the subsequent sexes of the children are given in the table below. (Notice that in this case the only value in the table that is predetermined or fixed by the experimenter is the overall sample size. Row totals, column totals, and cell frequencies are all random.)

Pregnancy Zone	Sex		
	Male	Female	
Present	51	78	129 (random)
Absent	96	75	171 (random)
	147 (random)	153 (random)	300 (fixed)

Find the error rates of this diagnostic test.

Sol'n:

Note: The technique demonstrated above can be used to estimate conditional probabilities in settings other than the diagnostic tests. Caution: if all row and column totals are random, then any conditional probability can be approximated. If not, then the only probabilities that can be approximated are those for which the sample sizes are fixed by the researcher. Consider the next example.

Ex. Suppose that a new home-pregnancy test has been developed. An experiment is conducted to approximate the false-positive and false-negative rates of the test. Five women who are known to be pregnant and 10 women who are not pregnant are selected to participate in the study. The new test is used on each and the results are given below (data are fictitious).

Test Result	True state		
	Not Pregnant	Pregnant	
Pregnant	5	1	6 (random)
Not pregnant	5	5	9 (random)
	10 (fixed)	5 (fixed)	15 (fixed)

Some conditional probabilities can be reliably approximated from these data whereas others cannot. For instance,

P(woman is pregnant / test pregnant) cannot be approximated

But

P(test pregnant / woman is pregnant) can be approximated.

Specificity of a test = the probability that the test result will be negative given that the subject is a true negative

Sensitivity of a test = the probability that the test result will be positive given that the subject is a true positive

Ex. In a study of 300 pairs of twins, the twins were questioned as to whether they were identical. Then other factors such as ABO blood group, MN blood type, and Rh blood type were considered. On the basis of these traits, the twins were classified as identical (+) or nonidentical (-). The latter classification procedure is considered to be the true classification. The purpose of the study is to test the ability of the twins to self-classify. The results are shown below. All row and column totals are random.

Self-classification	True classification		
	Nonidentical (-)	Identical (+)	
+	12	54	
-	130	4	
			200

Approximate the false-positive and false-negative rates of the self-classification procedure. Approximate also the specificity and sensitivity of the test.

Sol'n:

References:

1. Ronald E. Walpole. Introduction to Statistics, 3rd Edition. Macmillan Publishing Co. Inc., 1982
2. William Mendenhall and Richard Sheaffer. Mathematical Statistics with Applications. Wadsworth Publishing Company, 1973
3. YA- Lun Chou. Probability and Statistics for Decision Making, 1973
4. Lyman Ott et al.. Statistics: A tool for the Social Sciences, 3rd Edition, Duxbury Press, 1983.
5. John Reinard, Communication Research Statistics, Sage Publications, 2006.