University of Jordan
Faculty of Medicine
Batch 2013-2019

Epidemiology & Biostatistics

● Slides ○ Sheet ○ Handout ○ other.................

number #2
Title: Probability
Dr. Sireen al-Khaldi
Done By:
Date:
Price:

Designed by Nadeen Al-Freihat

Introduction to Biostatistics

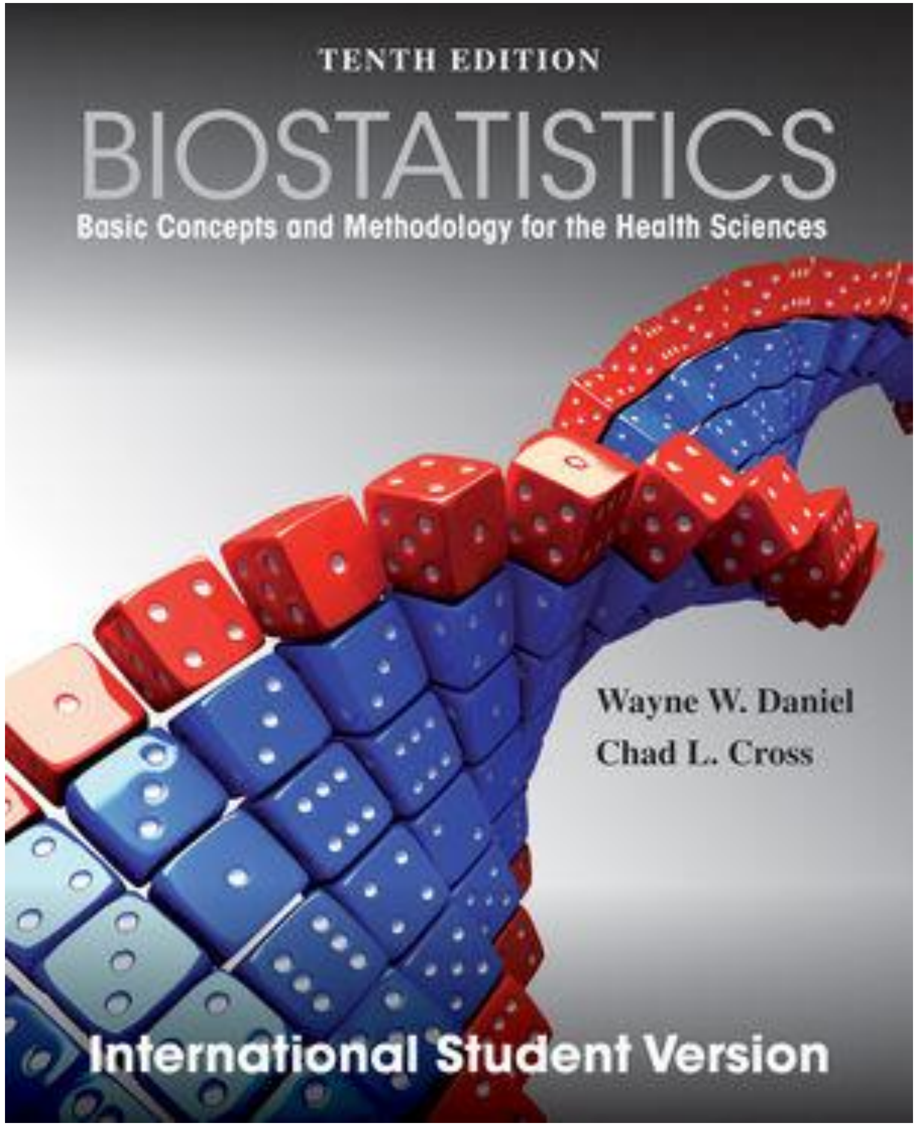
Probability

Second Semester 2014/2015

Text Book:

Basic Concepts and Methodology for the Health Sciences

By Wayne W. Daniel, 10 th edition



Dr. Sireen Alkhaldi, BDS, MPH, DrPH
Department of Family and Community Medicine
Faculty of Medicine
The University of Jordan

Chapter 3

Some Basic Probability Concepts

Learning Outcomes:

After studying this chapter, you will be able to:

1. Understand objective (classical, relative frequency), and subjective probability.
2. Understand the properties of probability and some probability rules.
3. Calculate the probability of an event.
4. Apply Baye's theorem to screening test results (sensitivity, specificity, and predictive value positive and negative)

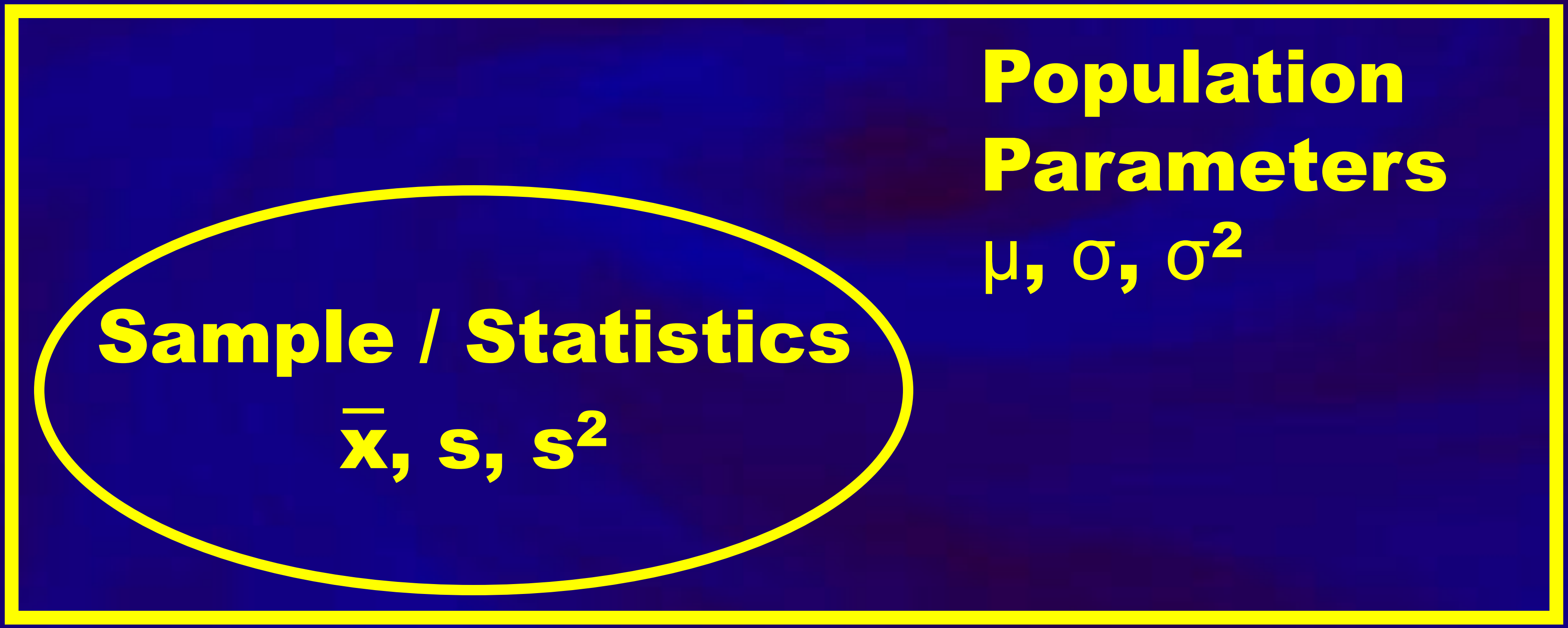
Introduction

The **theory of probability** is a branch of mathematics, but only its fundamental concepts will be discussed here.

This will provide the foundation for statistical inference (to reach a conclusion about a population from a sample drawn from that population).

The Big Picture

Populations and Samples



Introduction, continued

The concept of probability is frequently encountered in everyday communication.

For example:

- a physician may say that a patient has a 50-50 chance of surviving a certain operation.
- Another physician may say that she is 95 percent certain that a patient has a particular disease.
- A nurse may say that nine times out of ten, a client will break an appointment.

It is
all about
how
you
interpret
the
results!

"We'll only do 72% of it, since it's been reported that 28% of all surgery is unnecessary."

"Nine out of ten doctors think excessive drinking is bad for your health. Luckily mine is the tenth."

Introduction, continued

Those people have expressed probabilities mostly in terms of percentages (Probabilityx100).

- ✓ But, it is more convenient to express probabilities as fractions.
- ✓ Thus, we measure the probability of the occurrence of some event by a number between 0 and 1.
- ✓ The more likely the event, the closer the number is to one. An event that **can't occur** has a probability of zero, and an event that is **certain** to occur has a probability of one.

Two views of Probability

- **Objective Probability:**

1. Classical
2. Relative

- **Subjective Probability**

1. Classical Probability :

This theory was developed to solve the problems related to games of chance (rolling the dice or playing cards).

For Example:

If a **fair** six-sided die is rolled, the probability that a 1 will be observed is 1/6, and is the same for the other five faces.

If a card is picked from a **well-shuffled** deck of ordinary playing cards, the probability of picking a heart is 13/52.

If a fair six-sided die is tossed, the probability of an even numbered outcome (2, 4, 6) is ½. Three of the six equally likely outcomes have the trait (3/ 6 = ½)

1. Classical Probability

Definition:

If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a triat, E, the probability of the occurrence of event E is equal to m/ N [probability of E: P(E)= m/N]

2. **Relative Frequency Probability:**

Definition: If some process is repeated a large number of times, n, and if some resulting event E occurs m times , the relative frequency of occurrence of E , m/n will be approximately equal to probability of E.

Subjective Probability : (personalistic)

This concept does not rely on the repeatability of a process. It applies for events that can happen only once. It depends on personal judgement.

For Example : the probability that a cure for cancer will be discovered within the next 10 years.

Some important symbols

1.Equally likely outcomes: Are the outcomes that have the same chance of occurring.

2. A $\cap$B : Both A **and** B occur simultaneously (involves multiplication)

3. A U B : Either A **or** B occur, or they both occur (involves addition)

2.Mutually exclusive: Two events are mutually exclusive if they cannot occur simultaneously such that A $\cap$ B =Φ (events do not overlap)

3. The universal Set (S): The set of all possible outcomes.

4. The empty set Φ : Contain no elements.

5. The event ,E : is a set of outcomes in (S) which has a certain characteristic.

6. $\bar{A}$ or A' denotes the absence of A, that is occurrence of " not A".

Elementary Properties of Probability:

1. All events must have a probability greater than or equal to zero.

$P(E_i) \geq 0$, i= 1,2,3,…….n

2. The probability of all possible events should total to one (exhaustiveness)

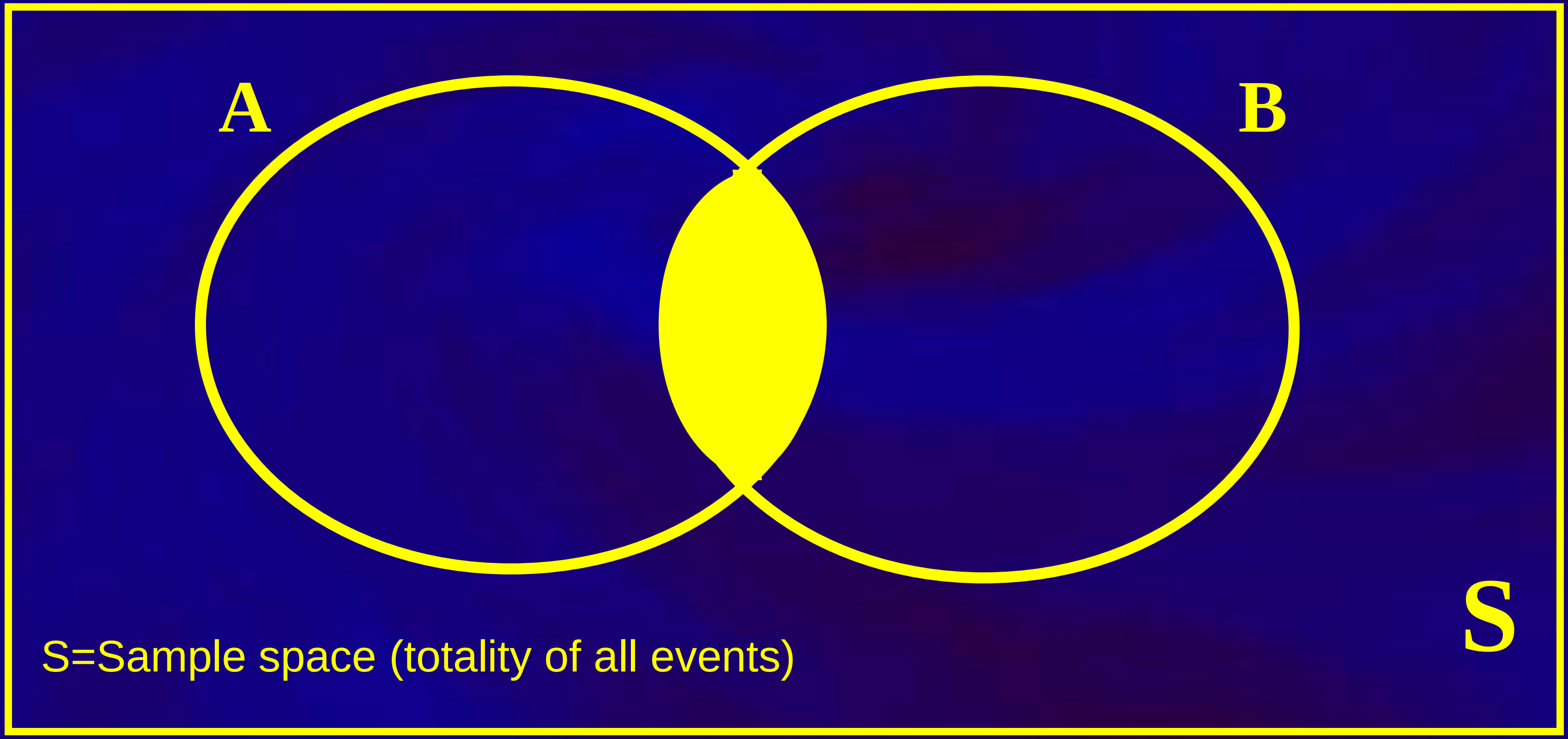
$P(E_1) + P(E_2) + \ldots\ldots + P(E_n) = 1$

3. Considering any two mutually exclusive events, the probability of the occurrence of either of them is equal to the sum of their individual probabilities.

$P(E_i + E_J) = P(E_i) + P(E_J)$ E_i ,E_J are mutually exclusive

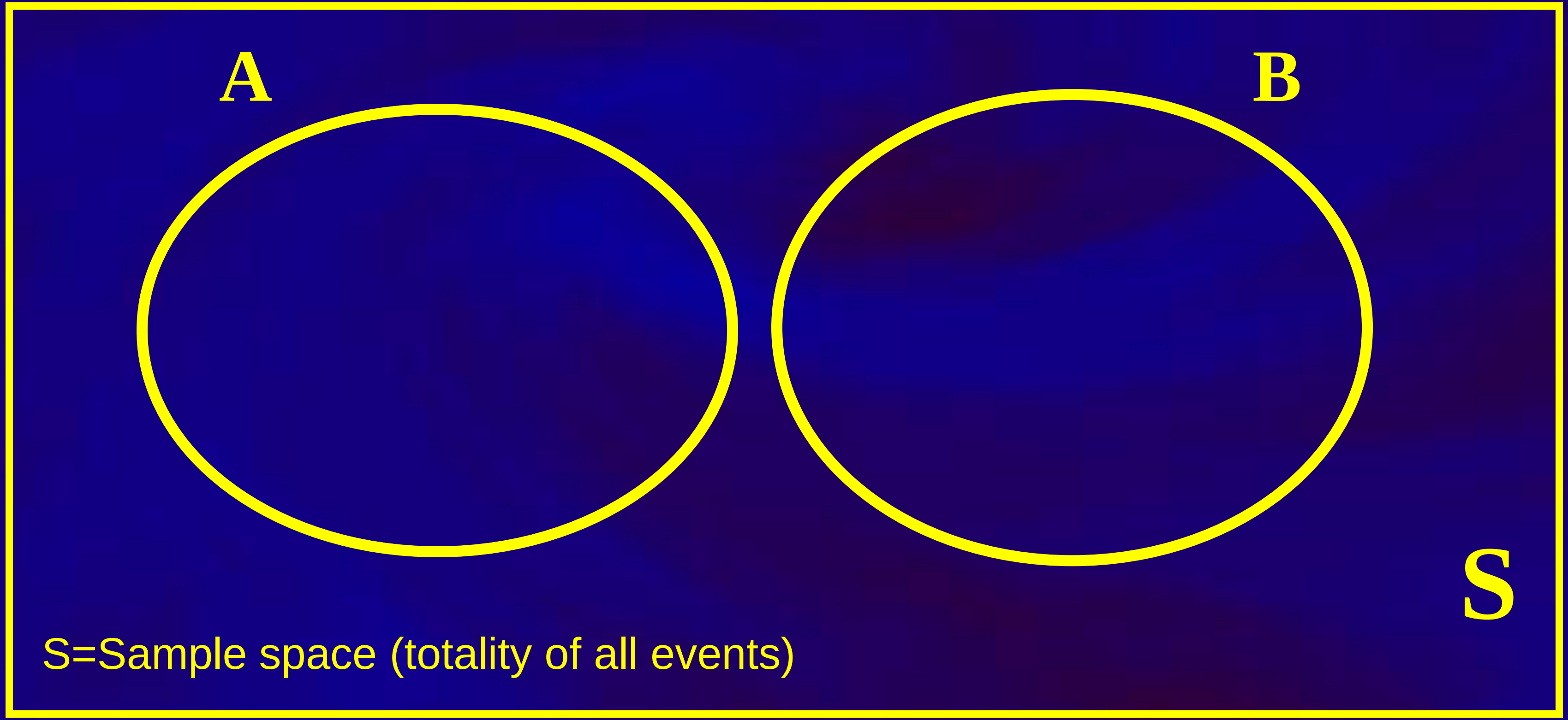
Intersection

- A∩B



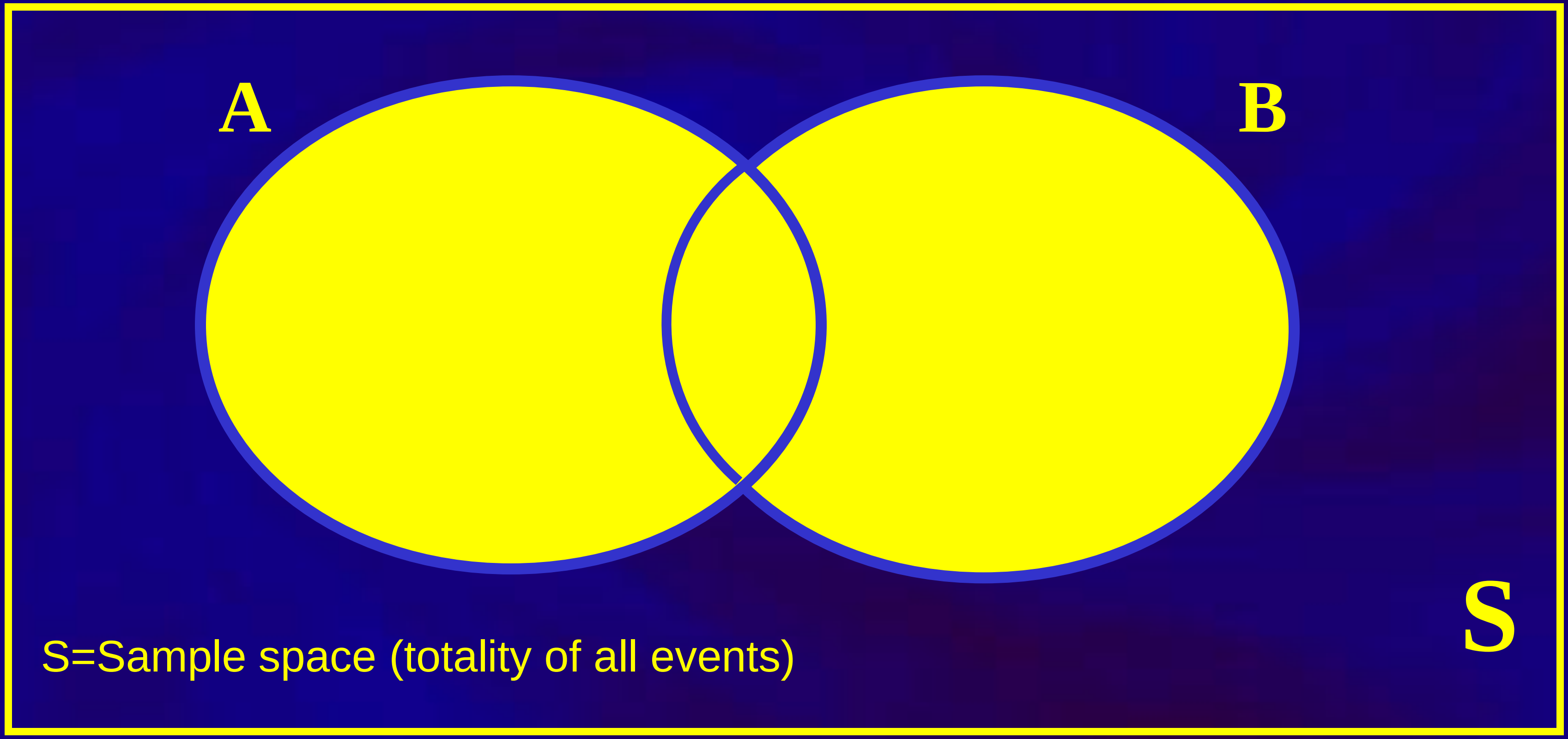
Mutually Exclusive

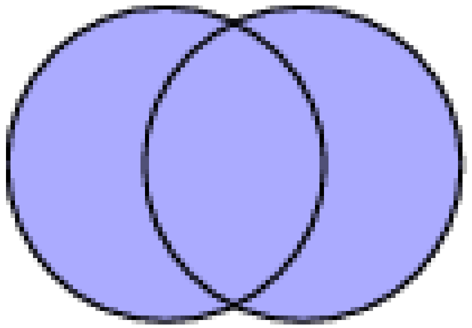
- **Implies no intersection**
- **Example: $(A \cap A^C) = \varnothing$ by definition**



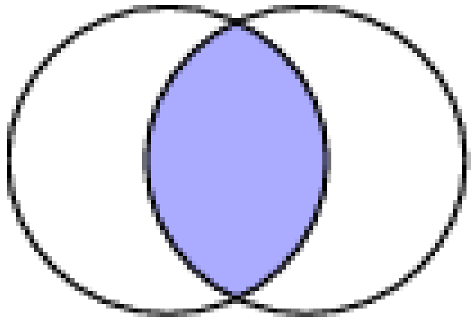
Union

- A∪B

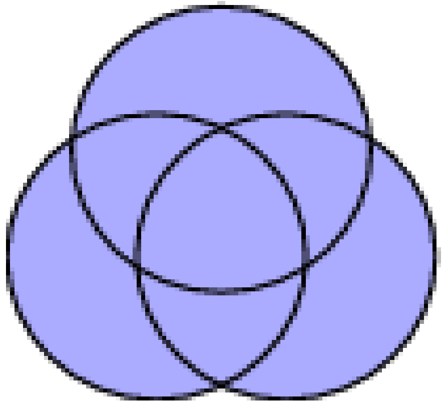




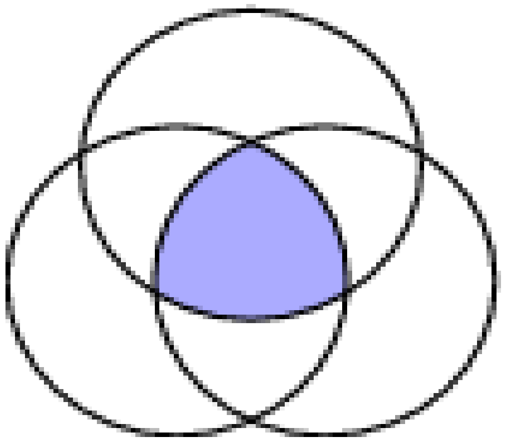
A U B “ A or B ”



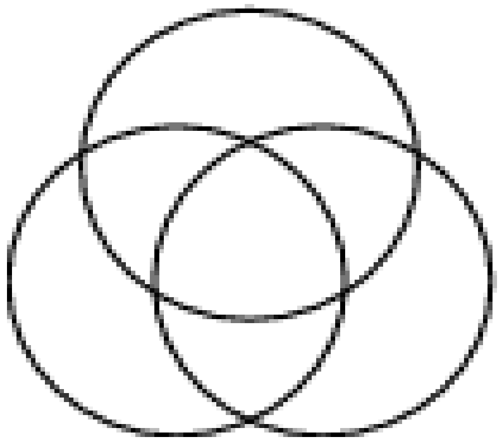
A ∩ B “ both A and B”
Intersection



A U B U C “ at least one”
A or B or C



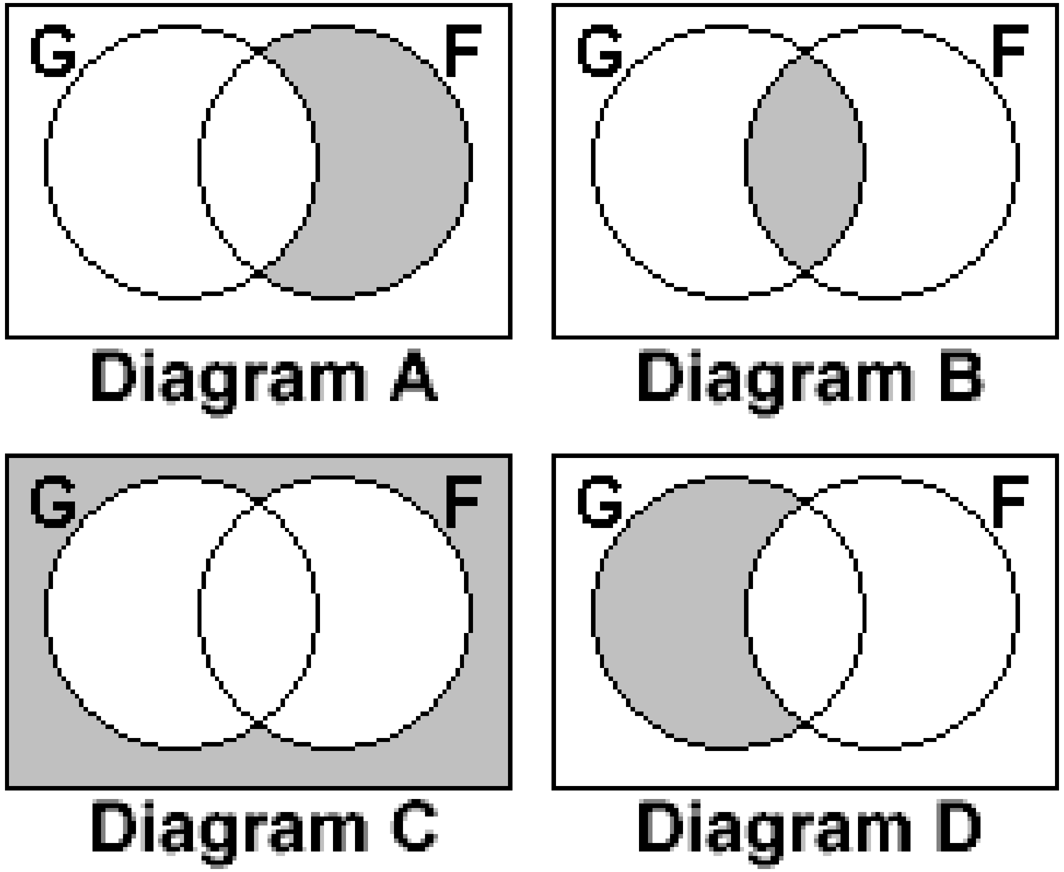
A ∩ B ∩ C “ All”
A and B and C



None

Exercise:

The diagrams below represent a class of children (boys and girls).
G is the set of girls and F is the set of children who like Healthy Food.



Girls who like Healthy Food .... B
Girls who dislike Healthy food .... D
Boys who like Healthy Food .... A
Boys who dislike Healthy Food C

Table 3.4.1 Frequency of family history of mood disorder by the age group among bipolar subjects

Family history of Mood Disorders	Early = 18 (E)	Later >18 (L)	Total
Negative(A)	28	35	63
Bipolar Disorder(B)	19	38	57
Unipolar (C)	41	44	85
Unipolar and Bipolar(D)	53	60	113
Total	141	177	318

If we pick a person at random from this sample, What is the probability that this person will be 18 years old or younger?

Answer the following questions:

Suppose we pick a person at random from this sample.

- ✓ The probability that this person will be 18-years old or younger?
- ✓ The probability that this person has family history of mood orders Unipolar(C)?
- ✓ The probability that this person has no family history of mood orders Unipolar($\overline{C}$)?
- ✓ The probability that this person is 18-years old or younger or has no family history of mood orders Negative (A)?
- ✓ The probability that this person is more than18-years old and has family history of mood orders Unipolar and Bipolar(D)?

Calculating the Probability of an Event

In the previous table:

- ✓ **The 318 subjects are the population**
- ✓ **Early and Late are mutually exclusive categories**
- ✓ **All persons are equally likely to be selected**

P (Early)= number of Early subjects/ total number of subjects

= 141 / 318 = 0.4434

Conditional Probability:

P(A\B) is the probability of A assuming that B has happened.

The probability of A given B.

P(A\B)= $\dfrac{P(A\cap B)}{P(B)}$, P(B)≠ 0

P(B\A)= $\dfrac{P(A\cap B)}{P(A)}$, P(A)≠ 0

Marginal probability (Unconditional probability)

One of marginal totals is used as nominator (141 / 318)

Conditional Probability

Exercises:

- Suppose we pick a person at random and find he is 18 years or younger (E),what is the probability that this person will be one who has no family history of mood disorders (A)? ……… A given E

The 141 Early subjects become denominator

The 28 Early subjects with (A) become the nominator

P(A\E) = 28/141= 0.1986

- suppose we pick a person at random and find he has family history of mood (D) what is the probability that this person will be 18 years or younger (E)? …E given D

P (E \ D) = 53/ 113= 0.469

	Early	Later	total
A	28	35	63
B	19	38	57
C	41	44	85
D	53	60	113
total	141	177	318

Joint Probability :

If a subject is picked at random from a group of subjects possesses *two characteristics* at the same time, this is called joint probability. It can be calculated as follows:

Suppose we pick a person at random from the 318 subjects. Find the probability that he will be Early (E) and will be a person who has no family history of mood disorders (A).

The number of subjects who satisfy

both conditions is found first:

$P\ (E \cap A) = 28/318 = 0.0881$

	Early	Later	total
A	28	35	63
B	19	38	57
C	41	44	85
D	53	60	113
total	141	177	318

Multiplicative Rule:

A probability can be computed from other probabilities.

For any two events A and B

- $P(A \cap B) = P(B)\, P(A \backslash B)$ if $P\,(B) \neq 0$
- $P(A \cap B) = P(A)\, P(B \backslash A)$ if $P\,(A) \neq 0$

Where,

- $P(A)$: marginal probability of A.
- $P(B)$: marginal probability of B.
- $P(B \backslash A)$:The conditional probability.

From this equation you can find any one of the three probabilities if the other two are known. This leads to: $P\,(A \backslash B) = P\,(A \cap B) / P(B)$

Independent Events:

- If event B has occurred and the probability of A is not affected by the occurrence or nonoccurrence of B, we say that A and b are independent.

1- $P(A \cap B) = P(B)\,P(A)$
2- $P(A \setminus B) = P(A)$
3- $P(B \setminus A) = P(B)$

Multiplicative Rule:

Example In a certain high school class consisting of 60 girls and 40 boys, it is observed that 24 girls and 16 boys wear eyeglasses . If a student is picked at random from this class ,the probability that the student wears eyeglasses , P(E), is 40/100 or 0.4 .

- What is the probability that a student picked at random wears eyeglasses given that the student is a boy?

$P(E\backslash B) = P(E \cap B) / P(B) = (16/100) / (40/100) = 0.4$

- What is the probability of the joint occurrence

of the events of wearing eye glasses and being a boy?

$P(E \cap B) = P(B)\, P(E\backslash B) = (40/100) \times 0.4 = 0.16$

- If you know that E and B are independent events:

$P(E \cap B) = P(B)\, P(E) = (40/100) \times (40/100) = 0.16$

	Eyegla-sses E	E’	total
Boy (B)	16	24	40
Girl (B’)	24	36	60
total	40	60	100

The Addition Rule

The addition rule is: P(AUB) = P(A) + P(B) – P (A∩B)

(U is “union” or “or”).

Example: If we pick a person at random from the 318 in the table, what is the probability that this person will be Early age onset (E) OR have no family history of mood disorders (A)

P (E U A) = (141/ 318) + (63/ 318) – (28/318)

= 0.4434 + 0.1981 - 0.0881 = 0.5534

	Early	Later	total
A	28	35	63
B	19	38	57
C	41	44	85
D	53	60	113
total	141	177	318

If A and B are mutually exclusive (disjoint) ,then P (A∩B) = 0

Then , addition rule is P(A U B)= P(A) + P(B) .

Complementary Rule

$P(\overline{A}) = 1 - P(A)$ where, $\overline{A}$ = complement event (A and $\overline{A}$ are mutually exclusive)

Early onset and Late onset are complementary events because the sum equals 1 .... [$P(A) + P(\overline{A}) = 1$]

Example

Suppose that of 1200 admission to a general hospital during a certain period of time,750 are private admissions. If we designate these as a set A, then compute P(A) , $P(\overline{A})$.

Complementary Events

Solution:

If we designate the 750 as a set A, then

$\overline{A}$ = 1200 – 750 = 450.

P (A) = 750 / 1200 = 0.625

P($\overline{A}$) = 450 / 1200 = 0.375

Also see that: P($\overline{A}$) = 1 – P (A) = 0.375

0.375 = 1 - 0.625

0.375 = 0.375

Summary of some Probability Rules

- P (A or B)= P (A) + P (B) – P (A and B).
- P (A or B)= P (A) + P (B) if A and B are mutually exclusive.
- P (A and B)= P (A) . P (B \ A).
- P (A and B)= P (A) . P (B) if A and B are independent.

Baye’s Theorem
Screening Tests, Sensitivity, and Specificity

Probability laws and concepts are widely applied in health sciences specially for the evaluation of screening tests and diagnostic criteria.

How can we enhance the ability to correctly predict the presence or absence of a particular disease from knowledge of test results (positive or negative) and/ or from the status of presenting symptoms (present of absent)

- ✓ What is the likelihood of a positive or negative result?
- ✓ What is the likelihood of the presence or absence of a particular symptom with or without a particular disease?

In screening tests we must be aware that the result is not always right.

TABLE 3.5.1 Sample of n Subjects (Where n Is Large) Cross-Classified According to Disease Status and Screening Test Result

	Disease		
Test Result	Present (D)	Absent ($\bar{D}$)	Total
Positive (T)	a	b	$a+b$
Negative ($\bar{T}$)	c	d	$c+d$
Total	$a+c$	$b+d$	n

Baye's theoremScreening tests, Sensitivity and Specificity

Prevalence Rate = (a + c)/ N

A False Positive results when a test indicates a positive status when the true status is negative

A False Negative results when a test indicates a negative status when the true status is positive.

The sensitivity of the test (or symptom) is the probability of positive test result given the presence of the disease. It is equals

$P(T \setminus D) = a / (a + c)$

The specificity of the test (or symptom) is the probability of a negative test result given the absence of the disease. It equals

$P(\overline{T} \setminus \overline{D}) = d / (b + d)$

Baye's Theorem, Screening tests

The Predictive value positive of the screening test (or symptom):

the probability that a subject has the disease given that the subject has a positive screening test result

$$P(D^+|T^+) = a / (a + b) *$$

The Predictive value Negative of the screening test (or symptom):

the probability that a subject does not have the disease, given that the subject has a negative screening test result

$$P(D^-|T^-) = d / (c + d) *$$

** if the prevalence of disease in the general population is the same as the prevalence of disease observed in the study*

Example 1

A medical research team wished to evaluate a proposed screening test for Alzheimer's disease. The test was given to a random sample of 450 patients with Alzheimer's disease and an independent random sample of 500 patients without symptoms of the disease. The two samples were drawn from populations of subjects who were 65 years or older. The results are as follows.

Test Result	Yes (D)	No ($\overline{D}$)	Total
Positive (T)	436	5	441
Negative ($\overline{T}$)	14	495	509
Total	450	500	950

Example 1: Screening test

a) Compute the false positive?

P ($\bar{D}$ \ T) = 5 / 441 = 0.0113

b) Compute the false negative?

P (D\ $\bar{T}$) = 14 / 509 = 0.0275

c) Compute the sensitivity of the screening test. $P(T \mid D) = \frac{436}{450} = 0.9689$

d) Compute the specificity of the screening test. $P(\bar{T} \mid \bar{D}) = \frac{495}{500} = 0.99$

e) Compute the Predictive value +. P V + = (D \ T)= 436/ 441 = 0.99

d) Compute the Predictive value _. P V - = (D'\ T')= 495/ 509 = 0.97

Exercise

Suppose that a certain ophthalmic trait is associated with eye color. Three hundred randomly selected individuals are studied with results as in the table.

Using these data, find:

1. P (trait)
2. P (blue eyes and trait)
3. P (brown eyes/ trait)

	Eye Color			Total
	Blue	Brown	Other	
Trait T	70	30	20	120
T’	20	110	50	180
Total	90	140	70	300

Answers:

	Eye Color			Total
	Blue	Brown	Other	
Trait T	70	30	20	120
T’	20	110	50	180
Total	90	140	70	300

1. P (trait): 120/ 300 =

2. P (blue eyes and trait)= P(blue\t) P(T)=

=(120/300)/ (70 / 300) = 70/300

3. P (brown eyes\ trait) = P (brown ∩ Trait)/ P (T)

= (30/300)/ (120/300)= 30/120 = 0.1

Probabilities… Example

Age (years)	Weight (lb) 130–149	150–169	170–189	≧ 190	Marginal Total
30–39	10	20	20	40	90
40–49	10	15	50	70	145
50–59	5	15	50	40	110
60–69	5	10	15	25	55
Marginal Totals	30	60	135	175	**400**

Using the table above, find for an individual the probability that he:

a. Is in the age interval 40–49.
b. Is in the age interval 40–49 and weighs 170–189 lb.
c. Is in the age interval 40–49 or 60–69.
d. Is in the age interval 40–49 or 60–69 and weighs 150–169 lb.
e. Is in the age interval 40–49 given that he weighs 150–169 lb.
f. Weighs less than 170 lb.
g. Weighs less than 170 lb and is less than 50 years old.
h. Weighs less than 170 lb given that he is less than 50 years old.

Solution

- P (is in age interval 40-49)= 145/ 400
- P (is in age interval 40-49 and weighs 70-189 lb.)= 50/400= 0.125
- P (is in age interval 40-49 or 60-69)=(145/400)+(55/400)= .36+ .138
- P (is in age interval 40-49 or 60-69 and weighs 150-169 lb.)=

  = 14/400 + 10/400= .035+.025= 0.060
- P(is in age interval 40-49 given that he weighs150-169 lb.)= 15/60=.25
- P (weighs less than 170 lb.)= (30 + 60)/400= ....
- P (weighs less than 170 lb. and is < 50 y old)= (10+20+10+15)/400= ...
- P (weighs less than 170 lb. given that he is <50 y old)=

  = (10+20+10+15) /90+145 = 55/235 =0.234

Screening test example ....

An individual is selected at random from a convalescent home in which 30 percent have a particular disease and is given a screening test to detect the presence of the disease. Let D denote the event that the person selected has the disease and let S indicate a positive result on the screening test. The probability of a positive screening test result given that the person selected has the disease is 0.9. The corresponding probability for a nondiseased person is 0.2. What is the probability that a person has the disease given that the screening test is positive? That is, find $P(D \mid S)$.

Solution:

Prevalence =30/100= 30 with disease

Sensitivity = 0.9 = 30 x 0.9 = 27 (S\ D)

P (S\ D')= 0.2 x 70 = 14

Predictive value + =P(D\S) = 27/ 41= 0.66

	D	D'	
S	27	14	41
S'	3	56	59
	30	70	100