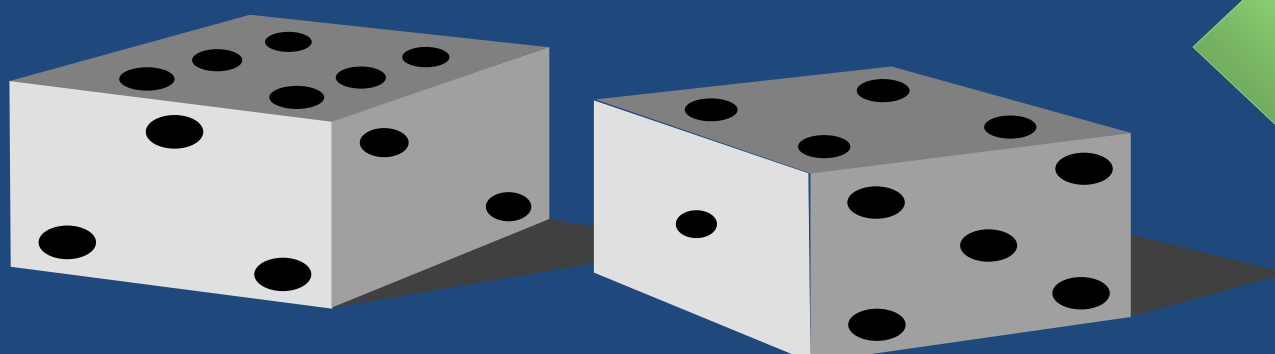
CHAPTER 3

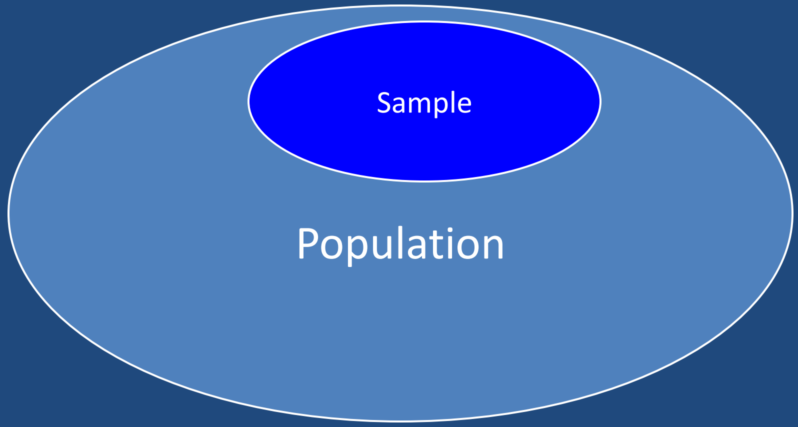
PROBABILITY: THE BASIS OF THE STATISTICAL INFERENCE



- **Key Words:**
  - Probability, Sample space, Event
  - Objective probability
  - Subjective probability
  - Equally likely
  - Mutually exclusive (Disjoint), Multiplicative rule
  - Conditional probability, Independent events
  - Total probability
  - Bayes theorem
  - Sensitivity, Specificity
  - Predictive value positive, Predictive value negative

3.1 Introduction

- Why – probability is the foundation of statistical inference
  - Methods needed to infer the characteristics of the population from which a sample was drawn



- 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.
- Most people express probabilities in terms of percentages.

- But, it is more convenient to express probabilities as fractions. Thus, we may 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.

- An Experiment is any act or planned process of data collection. It consists of a number of trials (replications) under the same condition.
- An experiment repeated under essentially homogeneous and similar conditions results in an outcome, which is unique or not unique but may be one of the several possible outcomes. When the result is unique then the experiment is called a ‘Deterministic' experiment.

- An experiment whose outcome cannot be predicted in advance, but is one of the set of possible outcomes, is called a ‘Random’ experiment. or ‘Nondeterministic’ experiment.

- If we think an experiment as being performed repeatedly, each repetition is called a trial. We observe an outcome for each trial.

Sample space (outcome space): is the set of all possible outcomes of a random experiment, associated with the random experiment and is denoted by S or Ω.

- Examples:

1. In the experiment of tossing a coin,

   $\Omega = \{H, T\}$

2. In the experiment of tossing two coins simultaneously,

   $\Omega = \{HH, HT, TH, TT\}$

3. In the experiment of throwing a pair of dice,

Ω = {(1,1), (1,2), (1,3), (1,4), …, (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

4. In the experiment of Disease test

Ω = {positive test, negative test}

- Event

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

- Event
  - Result of an experiment or observation
  - An Event is a subset of the sample space
  - Occurs or does not occur
  - Denoted by uppercase letters e.g. A,B, C, … or E_1, E_2, E_3, …
  - we will want to know the probability that an event occurs, P(A)
  - E.g. a disease occurrence, an extreme laboratory value

- **The universal Set** (Ω or S): The set all possible outcomes.

- **The empty set Φ** : Contain no elements.

- Two events, say A and B are said to be mutually exclusive (disjoint) if they cannot occur simultaneously (or $A \cap B = \Phi$)

- The outcomes that have the same chance of occurring are called Equally likely outcomes

3.2 Two views of Probability: Objective and Subjective:

- **Objective Probability**
  - –Classical Probability
  - –Relative Frequency Probability
- **Subjective (Personalistic) Probability**

– **Classical Probability** : If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a trait, E, the probability of the occurrence of event E is equal to m/ N .

Example:

in the rolling of the die , each of the six sides is equally likely to be observed . So, the probability that a 2 will be observed is equal to 1/6.

– Relative Frequency Probability:

If some posses 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 . $P(E) = m/n$.

Example: *Probability of Male versus Female Births*

Long-run relative frequency of males born in KOB is about 0.512 (512 boys born per 1000 births)

Table provides results of simulation: the proportion is far from 0.512 over the first few weeks but in the *long run* settles down around 0.512.

Relative Frequency of Male Births over Time

Weeks of Watching	Total Births	Total Boys	Proportion of Boys
1	30	19	19/30=0.633
2	116	68	68/116=0.586
13	317	172	172/317=0.543
26	623	383	383/623=0.615
39	919	483	483/919=0.526
52	1237	639	639/1237=0.517

– **Subjective Probability** :
Probability measures the confidence that a particular individual has in the truth of a particular proposition.

This approach reflects a personal opinion or best guess about whether an outcome will occur.

Examples :

- The probability that a cure for cancer will be discovered within the next 10 years.
- There is a 40% chance of rain tomorrow.

3.3 Elementary Properties of Probability

1. Given some process (or experiment) with *n* mutually exclusive outcomes (events), $E_1, E_2, \ldots, E_n$, the probability of any event E_i is assigned a nonnegative number. i.e.,

$$P(E_i) \geq 0, \quad i = 1, 2, 3, \ldots, n$$

2. The Sum of the probabilities of mutually exclusive and exhaustive outcomes is equal to 1

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

The events $E_1, E_2, \ldots, E_n$, are mutually exclusive and exhaustive if the do not overlap and their union is the sample space S.

Rules of Probability

1- $P(\Omega) = 1$, Ω is called the sure event

2- $P(\phi) = 0$, ϕ is called the null (impossible) event

3- $P(A^C)= 1-P(A)$, A^C : complement of A

4- If A and B are mutually exclusive (disjoint) events, then $P(A \cap B) = 0$, Then , addition rule is

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

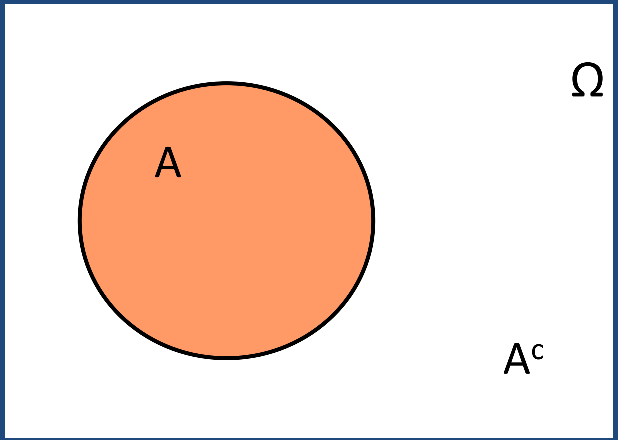
5- If A and B are any two events, then

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

6- If the events $A_1, A_2, \ldots, A_n$, are mutually exclusive then $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots P(A_n)$

3.4 Calculating The Probability of An Event

- Complement of an event, $\bar{A}$ or A^C (read Not A or A complement)
  - E.g. the event that the person does not have malaria
  - $P(A)= 1-P(A^c)$
    - In epidemiology, we often write E for exposed and $\bar{E}$ for not exposed
- $P(\Omega) = P(A) + P(A^c) = 1$



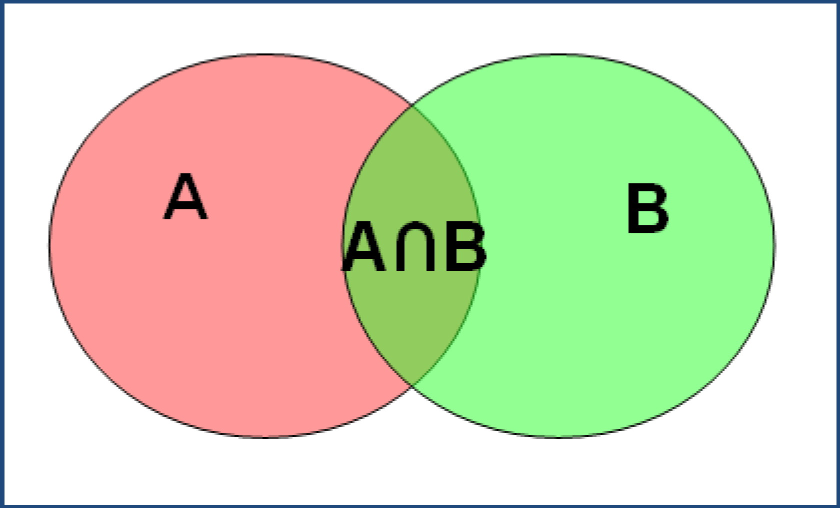
Complement example

Probability that someone has extremely drug resistant (XDR TB) versus they do not

$$P(\text{XDR TB+}) + P(\text{XDR TB-}) = 1$$

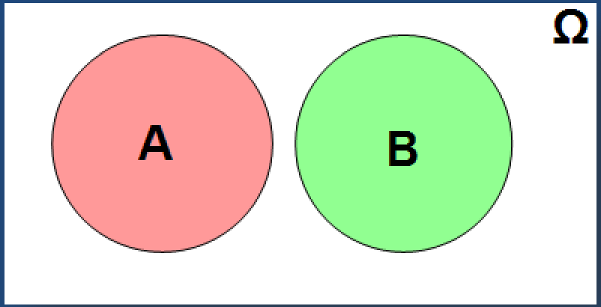
- **The intersection** of 2 events is written A ∩ B
- The intersection is when both A and B occur
  - E.g. The event that a person has both malaria and pulmonary tuberculosis

The probability that both occur is written P(A ∩ B)



- Two events are **mutually exclusive** if they cannot occur together ($A \cap B = \phi$)
  - In English: for mutually exclusive events, the probability of A or B occurring is the sum of their individual probabilities; both cannot occur together so $P(A \cap B) = 0$
  - In probability lexicon:

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



  - E.g.
    - Being pregnant and not pregnant
    - You cannot be both

- Example: (HCV: Hepatitis C Virus)

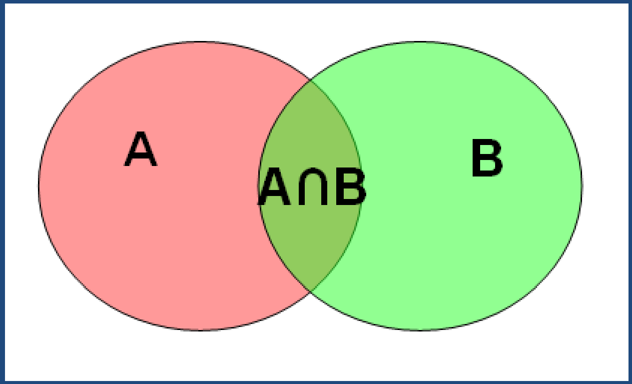
P(HCV genotype 1) in the US = 0.7

P(HCV genotype 2) in the US = 0.15

P(HCV genotype 3,4,6) = 0.15

→P(HCV genotype 1 or 2) = 0.85

- **The union** of 2 events is written A U B
- The union is if either A or B or both occur
  - E.g. The event that a person has either malaria or tuberculosis or both
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - The probability of A or B is the sum of their individual probabilities minus the probability of their intersection



Example

Let

- A = the event that an individual is exposed to high levels of carbon monoxide
- B= the event that an individual is exposed to high levels of nitrogen dioxide
  - What is the event A ∩ B called? What is that in this example?
  - What is the event A U B called? What is it in this example?
  - What is the complement of A?
  - Are A and B mutually exclusive?

Example: solution

- $A \cap B$ is the intersection of A and B. It is the event that the person is exposed to both gases.
- A U B is the union of A and B. It is the event that the person is exposed to one or the other or both.
- A^c is the event that the person is not exposed to carbon monoxide.
- Are A and B mutually exclusive? Can they both occur? Yes. So NOT mutually exclusive.

Example 3.4.1 : Frequency of Family History of Mood Disorder by 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

Suppose we pick a person at random from this sample, what the probability that this person:

1. Will be 18-years old or younger? 141/318=0.4434
2. Has family history of mood orders Unipolar (C)? 85/318=0.2673
3. Has no family history of mood orders Unipolar (C^c)? 1-(85/318)=0.7327
4. Is 18-years old or younger <u>or</u> has no family history of mood orders Negative (A)? (141/318+255/318-113/3180=0.8899)
5. Is more than 18-years old <u>and</u> has family history of mood orders Unipolar and Bipolar(D)?

60/318=0.1887

Conditional probability

- The probability that an event A will occur given that event B has occurred
  - Notation: P(A|B)
  - Read: the probability of A given B
- **E**xample: Probability of a person becoming infected with malaria given that he/she uses a bed net at night
  - Event B is using a bed net
  - Event A is becoming infected with malaria

Conditional Probability

The probability of an event when partial knowledge about the outcome of an experiment is known, is called Conditional probability.

P(A|B) = The conditional probability that event A occurs, given that event B has occurred.

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}, \ P(B) > 0}$$

Example 3.4.2 Page 71

From previous example 3.4.1 Page 69 , answer

- suppose we pick a person at random and find that 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)? $P(A|E)=28/141=0.1986$

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

$P(E|D)=53/113=0.4690$

Example 3.4.3.Page 71: Calculating a joint Probability

- Suppose we pick a person at random from the 318 subjects. Find the probability that he will early (E) and has no family history of mood disorders (A). $P(E \cap A)=28/318=0.0881$

- Multiplicative rule of probability

$$P(A \cap B) = P(A)\, P(B|A),\ P(A)>0$$

$$P(A \cap B) = P(B)\, P(A|B),\ P(B)>0$$

- Example 3.4.4 Page 72:

we wish to compute the joint probability of Early age at onset (E) and a negative family history of mood disorders (A) from a knowledge of an appropriate marginal probability and an appropriate conditional probability.

$$P(E \cap A) = P(E)\, P(A|E) = (141/318)*(28/141)=28/318=0.088$$

- Read Examples 3.4.5 and 3.4.6 Page 72-73

Example: The probabilities that a randomly selected woman who gave birth in 1992 (USA birth statistics)

Age of mother	Probability
<15	0.003
15-19	0.124
20-24	0.263
25-29	0.290
30-34	0.220
35-39	0.085
40-44	0.014
45-49	0.001
Total	1.000

Probability that mother's age was ≤24

= 0.003 + 0.124 + 0.263 = 0.390

(What probability rule?)

- Given that a mother is under age 30, what is the probability that she is under age 20?

P(Mother's age<20 | Mother's age<30)

= P (Mother's age<20 and <30) / P(Mother's age <30)

= (0.003 + 0.124) / (0.003 + 0.124 + 0.263 + 0.290)

= 0.127 / 0.68 = 0.187

Independence

- If the occurrence of B does not depend on A,
  - then $P(B|A) = P(B)$
  - Example: Probability of becoming infected with malaria given that you wear a blue shirt = probability of becoming infected with malaria
  - Then the multiplicative rule is

$$P(A \cap B) = P(A)\,P(B)$$

  - Example: coin tosses – the probability of a heads on the 2^{nd} throw is independent of the outcome on the first throw

Independence

Note that independence ≠ mutual exclusivity!

- Mutual exclusivity
  - 2 events cannot both occur
  - $P(A \cap B) = 0$
- Independence
  - 2 events do not depend on each other
  - $P(B|A) = P(B)$
  - $P(A \cap B) = P(A)\,P(B)$

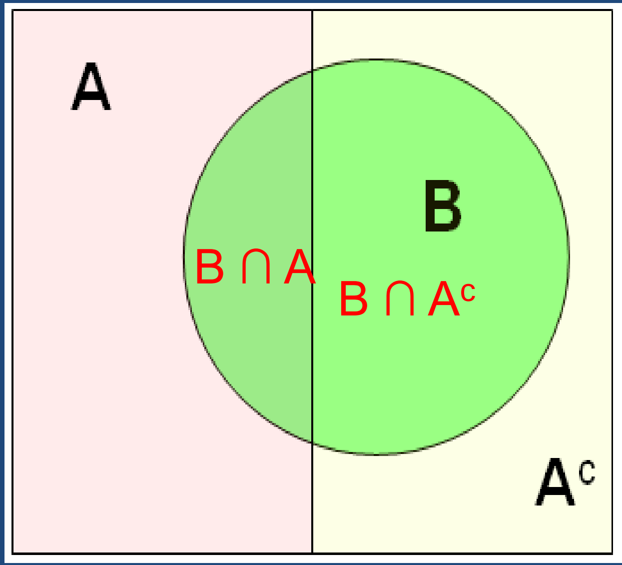
Example 3.4.7 Page 74

- 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 0.40 .
- What is the probability that a student picked at random wears eyeglasses (E) given that the student is a boy (B)? $P(E/B)= P(E \cap B)/P(B)= 0.40$
- What is the probability of the joint occurrence of the events of wearing eye glasses and being a boy?

$$P(E \cap B)=P(B)*P(E|B)= (0.40)*(0.40) = 0.16$$
$$= P(B)*P(E)$$

Law of Total (Marginal) Probability

$B = (B \cap A) \cup (B \cap A^c)$



The law of total probability:

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Bayes' theorem

Bayes' theorem allows you to use what you know about the conditional probability of one event on another to help you understand the inverse

$$P(A|\ B) = P(A \cap B) \ / \ P(B)$$

$$= P(\ B|A\ )\ P(A) \ / \ P(B)$$

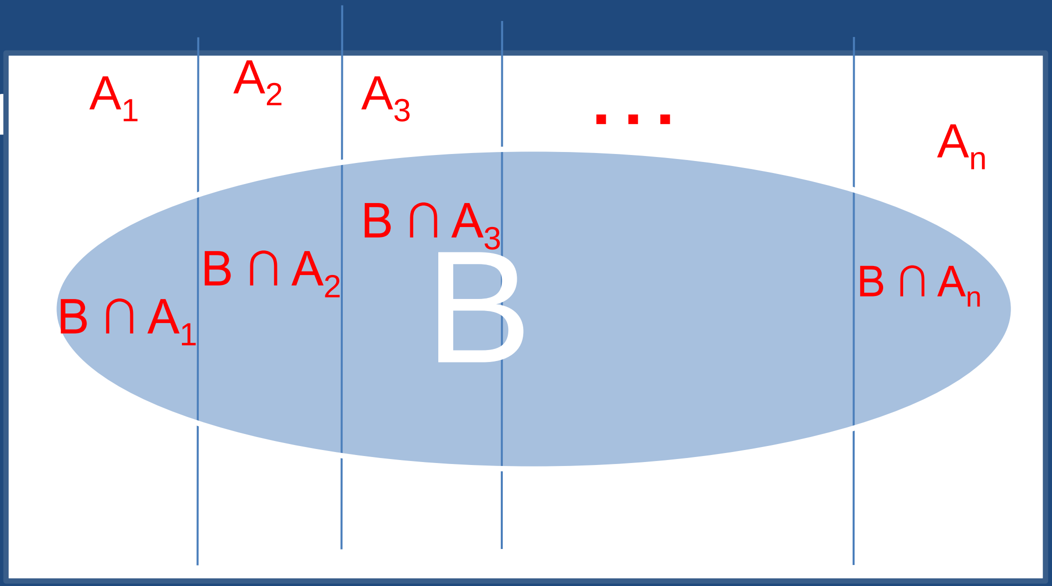
$$= P(\ B|A\ )\ P(A) \ / \ [P(B|A)P(A) + P(B|A^c)P(A^c)\ ]$$

Remember $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

Law of Total (Marginal) Probability

$A_1 \cup A_2 \cup \dots \cup A_n = \Omega$

$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1$



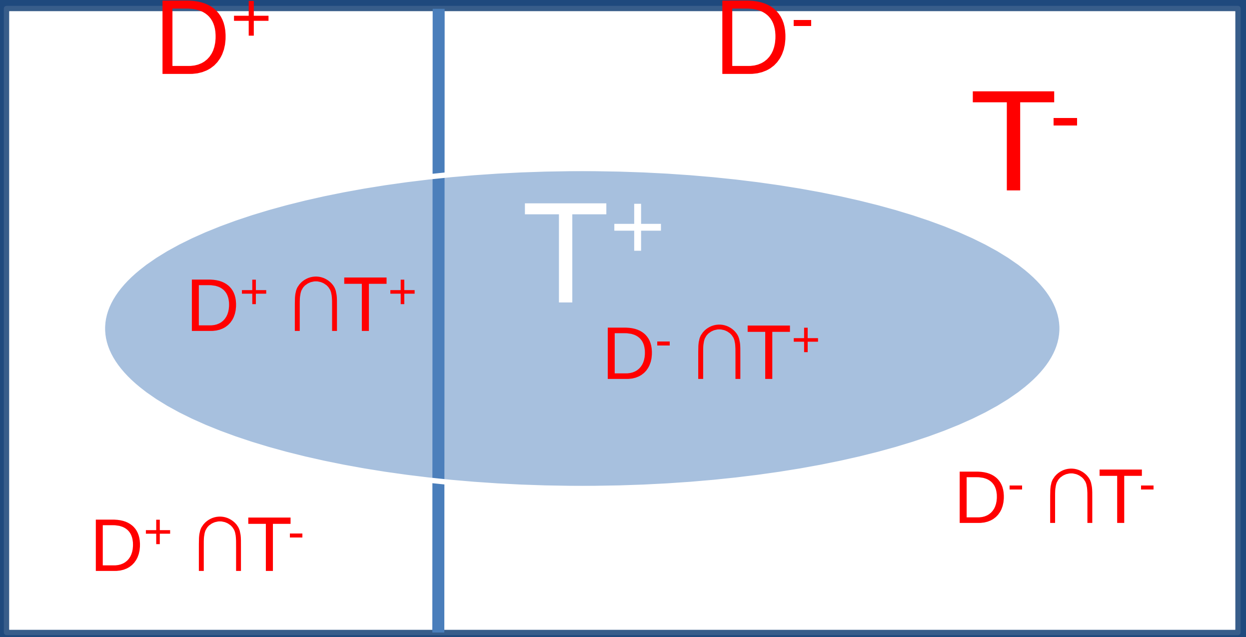
$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$

$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n)$

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n)$

Diagnostic Tests

D^+: Disease Present; D^-: Disease Absent



T^+: Test Positive; T^-: Test Negative

Diagnostic Tests

	Disease		
Test Result	Present (D^+)	Absent (D^-)	Total
Positive Test (T^+)	a **(TP)**	b **(FP)**	a + b
Negative Test (T^-)	c **(FN)**	d **(TN)**	c + d
	a + c	b + d	n

- False positives (FP) – the test is positive although the person does not have the disease
- False negatives (FN) – the test is negative even though the person has the disease

Diagnostic Tests

– True positives (TP) – the test is positive given the person has the disease

Sensitivity = The probability of this is $P(T^+|D^+)$

$$= P(T^+ \cap D^+)/P(D^+)$$

$$= TP/(TP+FN)=a/(a+c)$$

– True negatives (TN) – the test is negative given the person does not have the disease

Specificity =The probability of this is $P(T^-|D^-)$

$$= P(T^- \cap D^-)/P(D^-)$$

$$= TN/(FP+TN)=d/(b+d)$$

Diagnostic Tests

- Diagnostic test characteristics (sensitivity and specificity) are based on experiments in which the test is compared to a "gold standard"

Predictive Values

- Predictive value positive (PV^+): of a test is the probability that the subject has the disease given that the subject has a positive screening test
  - $PV^+ = P(D^+ \mid T^+)$

$$P(D^+ \mid T^+) = \frac{P(T^+ \mid D^+) * P(D^+)}{P(T^+ \mid D^+) * P(D^+) + P(T^+ \mid D^-) * P(D^-)}$$

$$P(D^+ \mid T^+) = \frac{(\text{sensitivity}) * P(D^+)}{(\text{sensitivity}) * P(D^+) + (1 - \text{specificity}) * (1 - P(D^+))}$$

- Predictive value negative (PV^-) of a test is the probability that a subject does not have the disease, given that the subject has a negative screening test
  - $PV^- = P(D^- \mid T^-)$

$$P(D^- \mid T^-) = \frac{P(T^- \mid D^-) * P(D^-)}{P(T^- \mid D^-) * P(D^-) + P(T^- \mid D^+) * P(D^+)}$$

$$P(D^- \mid T^-) = \frac{\text{specificity} * (1 - P(D))}{\text{specificity} * (1 - P(D)) + (1 - \text{sensitivity}) * P(D^+)}$$

Example: Sensitivity and specificity of breast cancer screening examination (HIP program)

	Breast Cancer		
Screening Test	Yes	No	Total
Positive	132	983	1115
Negative	45	63650	63695
Total	177	64633	64810

Sensitivity = a/(a+c) = 132/177 = 0.746

Specificity = d/(b+d) = 63650/64633 = 0.985

$PV^+ = P(D^+|T^+) = a/(a+b) = 132/1115 = 0.118$

$PV^- = P(D^-|T^-) = d/(c+d) = 63650/63695 = 0.999$

Example 3.5.1 Page 82

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.

	Alzheimer's Diagnosis		
Test Result	**Yes (D^+)**	**No (D^-)**	**Total**
Positive (T^+)	436	5	441
Negative (T^-)	14	495	509
Total	450	500	950

In the context of this example

a) What is a false positive?

A false positive is when the test indicates a positive result when the person does not have the disease

b) What is the false negative?

A false negative is when a test indicates a negative result when the person has the disease.

c) Compute the sensitivity of the symptom.

$P(T^+|D^+) = 436/450 = 0.9689$

d) Compute the specificity of the symptom.

$P(T^-|D^-) = 495/500 = 0.99$

e) Suppose it is known that the rate of the disease in the general population is 11.3%. What is the predictive value positive of the symptom and the predictive value negative of the symptom

The predictive value positive of the symptom is calculated as

$$P(D^+ \mid T^+) = \frac{(\text{sensitivity}) * P(D^+)}{(\text{sensitivity}) * P(D^+) + (1 - \text{specificity}) * (1 - P(D^+))}$$

$$= \frac{(0.9689)(0.113)}{(0.9689)(0.113) + (.01)(1 - 0.113)} = 0.925$$

The predictive value negative of the symptom is calculated as

$$P(D^- \mid T^-) = \frac{\text{specificity}*(1-P(D))}{\text{specificity}*(1-P(D)) + (1-\text{sensitivity})*P(D^+)}$$

$$= \frac{(0.99)(0.887)}{(0.99)(0.887) + (0.0311)(0.113)} = 0.996$$

Prevalence and Incidence

- Prevalence is the probability of having the disease or condition at a given point in time regardless of the duration
- Incidence is the probability that someone without the disease or condition will contract it during a specified period of time