Some Basic Probability Concepts

Objectives:

After studying this chapter, the student will:

1. understand classical, relative frequency, and subjective probability.
2. understand the properties of probability and selected probability rules.
3. be able to calculate the probability of an event.
4. be able to apply Bayes' theorem when calculating screening test results.

Introduction:

- The concept of probability is not foreign to health workers and is frequently encountered in everyday communication.
    - For example, we may hear a physician 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 disease.
    - The probability that a sick patient who receives a new medical treatment will survive for five or more years.
    - The probability it will rain tomorrow is 85%.

- Knowing the probability of these outcomes helps us make decisions.
- In this chapter, we assume that the population is known and calculate the chances of obtaining various samples from the population.
- Thus, the probability is the reverse of statistics; In probability, we use the population information to infer the probable nature of the sample.
- When we say an event is subject to chance, we mean that the outcome is in doubt and there are at least two possible outcomes.

- A **random experiment** is an experiment for which the outcome is unknown with certainty.
- The set of all possible outcomes of a random experiment is called a **sample space**.
- Any subset of the sample space is called an **event**.
- Probability is a mathematical construction that determines the likelihood of occurrence of events that are subject to chance.
- Probability is a numerical measure between 0 and 1 that describes the likelihood that an event will occur. Probabilities closer to 1 indicate that the event is more likely to occur. Probabilities closer to 0 indicate that the event is less likely to occur.
- The probabilities assigned to events must satisfy these requirements:

  1. The probability of any event must be nonnegative.
  2. The probability of the entire sample space must be 1.
  3. For two disjoint events A and B, the probability of the union of A and B is equal to the sum of the probabilities of A and B; that is,

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

- Two events are said to be disjoint (mutually exclusive) if $A \cap B = \emptyset$ (or $P(A \cap B) = 0$).

Three Conceptual Approaches to Probability:

There are three ways to assign probabilities to events.

1. Classical Probability:

- The outcomes of a random experiment that have the same chance of occurring are called equally likely outcomes.
- The classical probability rule is applied to compute the probabilities of events for an experiment in which all outcomes are equally likely.
- If an event can occur in $\mathbf{N}$ mutually exclusive and equally likely ways, and if $\mathbf{m}$ of these possess a trait $\mathbf{E}$, the probability of the occurrence of $\mathbf{E}$ is equal to $\frac{\mathbf{m}}{\mathbf{N}}$.

- If we read $\mathbf{P(E)}$ as "the probability of $\mathbf{E}$," we may express this definition as

$$P(E) = \frac{m}{N}$$

Example:

Consider the experiment of tossing a coin twice.

1. List the experimental outcomes (sample space).
2. What is the probability of obtaining one head and one tail?
3. What is the probability of obtaining at least one head?

Solution:

Example:

Find the probability of obtaining an even number in one roll o f a die.

Solution:

Example:

Human eye color is controlled by a single pair of genes (one from the father and one from the mother) called a genotype. Brown eye color, B, is dominant over blue eye color, L. Therefore, in the genotype BL, consisting of one brown gene B and one blue gene L, the brown gene dominates. A person with a BL genotype has brown eyes. If both parents have brown eyes and have genotype BL:

1. What is the probability that their child will have blue eyes?

2. What is the probability the child will have brown eyes?

Solution:

2. Relative Frequency Probability (Empirical Probability):

- Probabilities are assigned on the basis of experimentation or historical data.
- If some process is repeated a large number of times, $\mathbf{n}$, and if some resulting event with the characteristic $\mathbf{E}$ occurs $\mathbf{m}$ times, the relative frequency of occurrence of $\mathbf{E}$, $\frac{\mathbf{m}}{\mathbf{n}}$, will be approximately equal to the probability of $\mathbf{E}$.
- To express this definition in compact form, we write

  $$P(E) = \frac{m}{n}.$$

-

- We must keep in mind, however, that, strictly speaking, $\frac{m}{n}$ is only an estimate of $P(E)$.

Example:

In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood, and 2 had type AB blood. Set up a frequency distribution and find the probability a person has type O blood.

Solution:

3. Subjective Probability

- In the subjective approach, we define probability as the degree of belief that we hold in the occurrence of an event. Thus, judgment is used as the basis for assigning probabilities.

- The use of the subjective approach is usually limited to experiments that are unrepeatable.

Examples

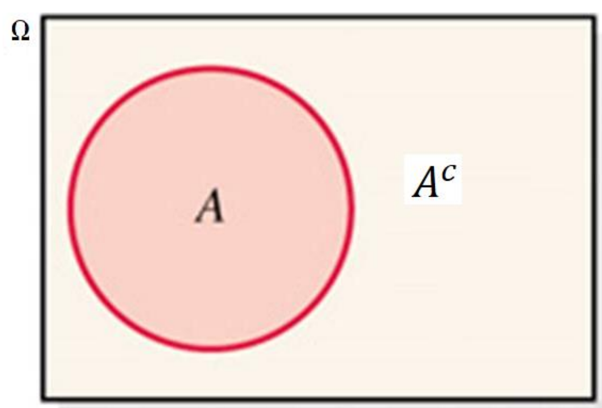
1. The probability of rain in the next 24 hours is very high.

2. Given a patient's health and extent of injuries, a doctor may feel that the patient has a 70% chance of a full recovery.

Calculating the Probability of an Event:

Complementary Event:

- The complement of an event A is the set of outcomes in the sample space that are not included in the outcomes of event A. The complement of A is denoted by A^c or $\overline{A}$.



- The rule for complementary events is:

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

Example:

If the probability that a person lives in an industrialized country of the world is 0.2, find the probability that a person does not live in an industrialized country.

Solution:

Example:

Suppose we have 5 medications which can be used to treat a particular case and 2 of these medications have undesirable side effects. If one of these medications is randomly chosen,

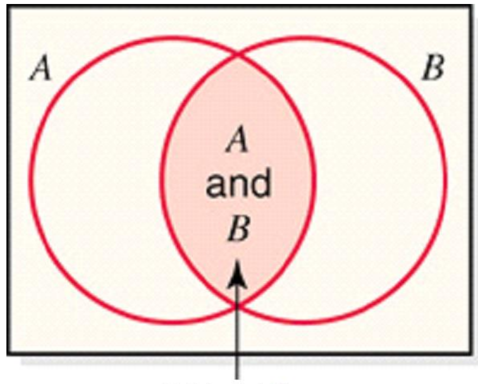
1. What is the probability that the medication will have an undesirable side effect?

2. What is the probability that the chosen medication will not have an undesirable side effect?

Solution:

Joint Probability:

- Sometimes we want to find the probability that a subject picked at random from a group of subjects possesses two characteristics at the same time. Such a probability is referred to as a **joint probability**.



$P(A \cap B)$

Example:

In a hospital unit there are 8 nurses and 5 physicians; 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject is a nurse and a male.

Solution:

Marginal Probability:

- **Marginal probability** is the probability of a single event without consideration of any other event.

Example:

In the previous example, find the probability that the person is a nurse.

Solution:

Conditional Probability:

- The sample space of interest may be reduced by conditions not applicable to the total group. When probabilities are calculated with a subset of the total group as the denominator, the result is a conditional probability.

- **Conditional probability** is the probability that an event will occur given that another event has already occurred.

- If A and B are two events, then the conditional probability of A **given** B is written as $P(A|B)$.

Example:

In the previous example, find the probability that a male person is a nurse.

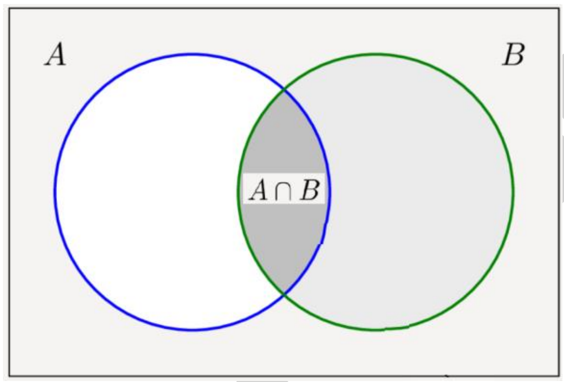
Solution:

Calculating Conditional Probability

- If A and B are two events, then,

  $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ given that $P(B) \neq 0$.

- The reasoning behind this definition is that if B has occurred, then only the "portion" of A that is contained in B, i.e., $A \cap B$, could occur; moreover, the original probability of $A \cap B$ must be recalculated to reflect the fact that the "new" sample space is B.



Example:

In the previous example, find the probability that a male person is a nurse.

Solution:

The Multiplication Rule:

- If A and B are two events, then the probability of A **and** B (joint) is written as $P(A \cap B)$.
- The multiplication rule is given by:

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

Example:

A statistics course has seven male and three female students. The professor wants to select two students at random to help him conduct a research project. What is the probability that the two students chosen are female?

Solution:

Independent Events:

- Two events, A and B, are two **independent events** if the occurrence of one does not affect the probability of the occurrence of the other. In other words, A and B are ind*ependent events* if either $P(A|B) = A$ or $P(B|A) = B$.
- The multiplication rule for independent events is given by:

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

- The events that are not independent are said to be ***dependent***.
- Two events are either mutually exclusive or independent.
  - Mutually exclusive events are always dependent.
  - Independent events are never mutually exclusive.
- Dependent events may or may not be mutually exclusive.

Example:

Approximately 9% of men have a type of color blindness that prevents them from distinguishing between red and green. If 3 men are selected at random, find the probability that all of them will have this type of red-green color blindness.

Solution:

The Addition Rule:

- Given two events A and B, the probability that event A, or event B, or both occur may be written as

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

Example:

In a group of 2500 persons, 1400 are female, 600 are vegetarian, and 400 are female and vegetarian. What is the probability that a randomly selected person from this group is a male or vegetarian?

Solution:

Example:

The following table shows physicians who smoked at some time classified according to their age and current frequency of smoking.

	Current frequency of smoking			
Age	**Daily**	**Occasionally**	**Not at all**	**Total**
20-29	31	9	7	**47**
30-39	110	30	49	**189**
40-49	29	21	29	**79**
50+	6	0	18	**24**
Total	**176**	**60**	**103**	**339**

A physician is selected at random, find the probability that

1. The physician is from 40-49 years old.
2. The physician is occasionally smoke.
3. The physician is from 40-49 years old and daily smoke.
4. The physician is from 30-39 years old or not at all smoke.
5. The physician is not 50 years old or older.
6. The physician is occasionally smoke given that he is from 40-49 years old group.

Solution:

Exercises:

Q1. Classify each statement as an example of classical probability, empirical probability, or subjective probability.

1. The probability that you will married by age 30 is 0.5.
2. The probability that a voter chosen at random will vote Republican is 0.45.
3. The probability of winning a 1000-ticket raffle with one ticket is 0.001.

Q2. Hospital records indicated that knee replacement patients stayed in the hospital for the number of days shown in the distribution.

Number of days stayed	Frequency
3	15
4	32
5	56
6	19
7	5
Total	**127**

Find these probabilities.

1. A patient stayed exactly 5 days.
2. A patient stayed at most 4 days.
3. A patient stayed less than 6 days.
4. A patient stayed at least 5 days.

Q3. A recent survey asked 100 people if they thought women in the armed forces should be permitted to participate in combat. The results of the survey are shown.

Gender	Yes	No	Total
Male	32	18	50
Female	8	42	50
Total	**40**	**60**	**100**

Find these probabilities.

1. The respondent answered yes, given that the respondent was a female.

2. The respondent was a male, given that the respondent answered no.

Q4. Suppose a sample of men are categorized by weight: "reasonable weight" or "overweight," and their blood pressures by "normal" or "high." Suppose that we have the following table:

	Blood Pressure		
Weight	**Normal (N)**	**High (H)**	**Total**
Reasonable (R)	0.6	0.1	**0.7**
Overweight (O)	0.2	0.1	**0.3**
Total	**0.8**	**0.2**	**1.0**

1. Find the probability that a person with reasonable weight has high blood pressure.

2. What is the probability that a randomly selected person is overweight and has high blood pressure?

Q5. A recent survey asked 100 people if they thought women in the armed forces should be permitted to participate in combat. The results of the survey are shown.

Gender	**Yes**	**No**	**Total**
Male	32	18	**50**
Female	8	42	**50**
Total	**40**	**60**	**100**

Find these probabilities.

1. The respondent answered "Yes", given that the respondent was a "female".

2. The respondent was a "male", given that the respondent answered "No".
3. The respondent is a "male" or the respondent's answer is "No".
4. Are events "Male" and "Yes" independent?
5. Are events "Male" and "Yes" mutually exclusive?

Q6. A recent study of 300 patients found that of 100 alcoholic patients, 87 had elevated cholesterol levels, and of 200 nonalcoholic patients, 43 had elevated cholesterol levels. If a patient is selected at random, find the probability that the patient is the following.

1. An alcoholic with elevated cholesterol level.
2. A nonalcoholic.
3. Anonalcoholic with nonelevated cholesterol Level.

Q7. In a pizza restaurant, 95% of the customers order pizza. If 65% of the customers order pizza and a salad, find the probability that a customer who orders pizza will also order a salad.

Q8. A medication is 75% effective against a bacterial infection. Find the probability that if 12 people take the medication, at least 1 person’s infection will not improve.

Q9. If we have a group of 10 staff members in a clinic with 3 physicians, 5 nurses, and 2 secretaries, suppose that the member that are married are 2 physicians, 3 nurses, and 2 secretaries. If we choose one staff member at random, find the probabilities that we choose

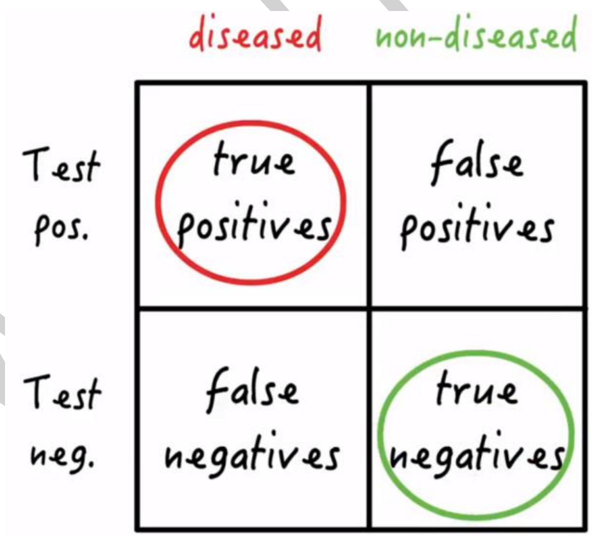
1. A married person.
2. An unmarried nurses.
3. An unmarried secretary.
4. A married physician or a nurse.
5. A physician and a secretary.

Sensitivity, Specificity, and Predictive Value:

- Consider the following table that gives the number of subjects with positive and negative tests, and with and without disease in a **cross-sectional** random sample from a target population.

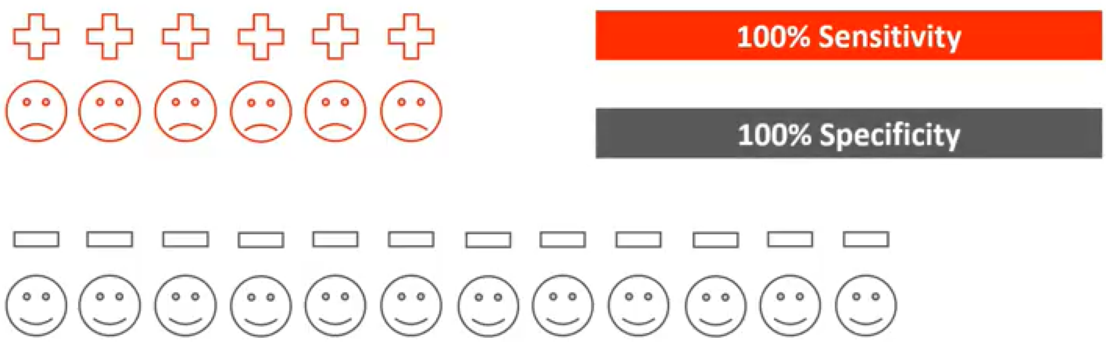
	Disease		
Test	**Present**	**Absent**	**Total**
Positive	a	b	$a+b$
Negative	c	d	$c+d$
Total	$\boldsymbol{a+c}$	$\boldsymbol{b+d}$	$\boldsymbol{n}$

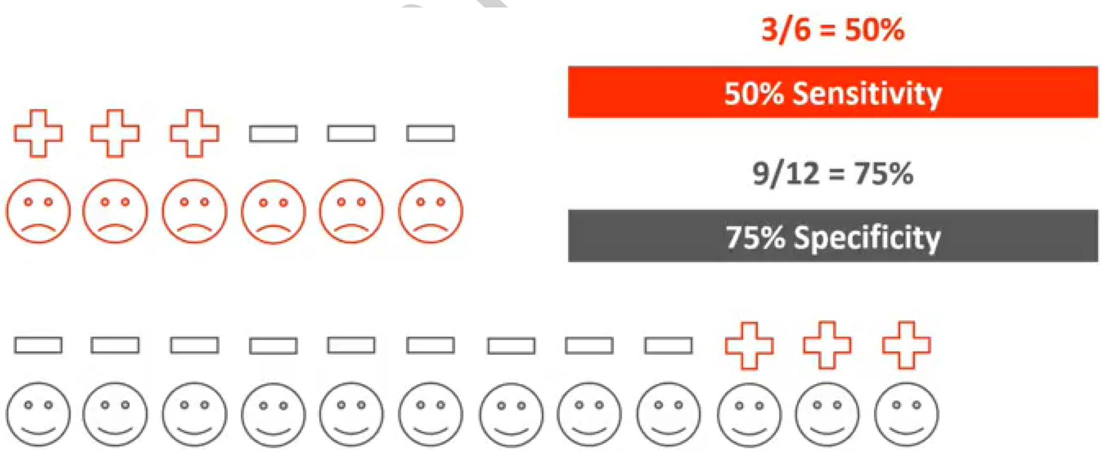
- **True positives** $= a$: Persons who have disease and are test positive.
- **True negatives** $= d$: Persons who do not have disease and are test negative.
- **False positives** $= b$: Persons without disease but with positive test.
- **False negatives** $= c$: Person with disease but with negative test.



- **Sensitivity** of the test $= \frac{a}{a+c}$: Ability of the test to be correctly positive among those who are known to have the disease (It is the probability of a positive test result given the presence of the disease). The corresponding errors are false negatives.
- **Specificity** of the test $= \frac{d}{b+d}$: Ability of the test to be correctly negative among those who are known to be without disease (It is the probability of a negative test result given the absence of the disease). The corresponding errors are false positives.

- Sensitivity and specificity are used to evaluate the **validity** of lab tests.
- A test is sensitive to the disease if it is positive for most people having the disease.
- A test is specific if it is positive for a small percentage of those without the disease.
- It is desirable that a test or screening procedure be highly sensitive and highly specific.





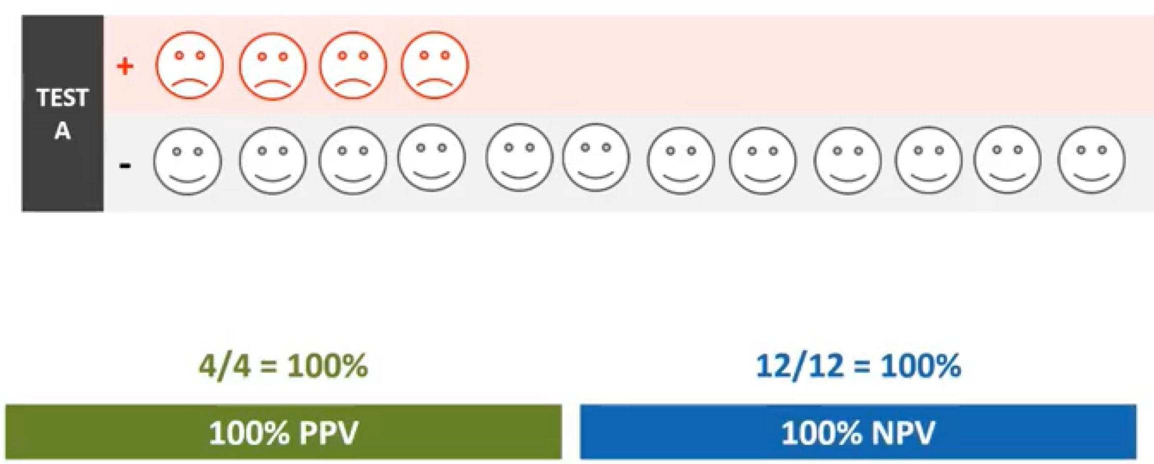
- **Predictive value positive** of the test $= \frac{a}{a+b}$: Ability of the test to correctly predict the presence of disease (It is the probability that a subject has the disease, given that the subject has a positive screening test result). Used to determining how to proceed after a patient gets a positive result.

- **Predictive value negative** of the test $= \frac{d}{c+d}$: Ability of the test to correctly predict the absence of disease (It is the probability that a subject does not have the disease, given that the subject has a negative screening test result). Used to determining how to proceed after a patient gets a negative result.

- A test with good positive predictivity is required for confirming the presence of disease, and a test with good negative predictivity to exclude the disease.

Population:

Test A:



Test B:

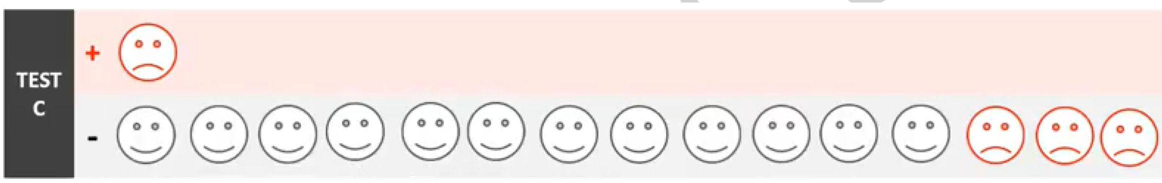
4/8 = 50%

50% PPV

8/8 = 100%

100% NPV

Test C:



1/1 = 100%

100% PPV

12/15 = 80%

80% NPV

Test D:

3/9 = 33.3%

33% PPV

6/7 = 85.7%

85.7% NPV

Example:

A cytological test was undertaken to screen women for cervical cancer. Consider a group of 24,103 women consisting of 379 women whose cervices are abnormal (to an extent sufficient to justify concern with respect to possible cancer) and 23,724 women whose cervices are acceptably healthy. A test was applied and results are tabulated in the following table.

	Disease		
Test	**Present**	**Absent**	**Total**
Positive	154	362	516
Negative	225	23362	23587
Total	379	23724	

1. Compute the sensitivity of the test.
2. Compute the specificity of the test.

Solution:

Bayes' Theorem:

- If prevalence of the disease in the target group is known, predictivity can be obtained by using sensitivity and specificity by the use of Bayes' Theorem.
- Sensitivity is denoted by $P(T|D)$.
- Specificity is denoted by $P(T^c|D^c)$.

- **The predictive value positive is given by**

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

$$= \frac{sens \cdot prev}{sens \cdot prev + (1 - spes) \cdot (1 - prev)}$$

- **The predictive value negative is given by**

$$P(D^c|T^c) = \frac{P(T^c|D^c)P(D^c)}{P(T^c|D^c)P(D^c) + P(T^c|D)P(D)}$$

$$= \frac{spes \cdot (1 - prev)}{spes \cdot (1 - prev) + (1 - sens) \cdot prev}$$

Example:

A medical research team wishes to assess the usefulness of a certain symptom (call it S) in the diagnosis of a particular disease. In a random sample of 775 patients with the disease, 744 reported having the symptom. In an independent random sample of 1380 subjects without the disease, 21 reported that they had the symptom.

1. Compute the sensitivity of the symptom.
2. Compute the specificity of the symptom.
3. Suppose it is known that the rate of the disease in the general population is .001. What is the predictive value positive of the symptom?
4. What is the predictive value negative of the symptom?

Solution:

Example:

The prevalence of a disease is 1 in 1000, and there is a test that can detect it with a sensitivity of 100% and specificity of 95%. What is the probability that a person has the disease, given a positive result on the test?

Solution:

Exercises:

Q1. Consider the data shown below on the use of x-ray as a screening test for Tuberculosis.

	Tuberculosis		
X-ray	No	Yes	Total
Negative	1739	8	1747
Positive	51	22	73
Total	1790	30	1820

1. In the context of this exercise, what is a false positive?
2. Calculate the sensitivity and specificity.
3. Calculate the predictive value positive when the prevalence is 0.2.

Q2. The sensitivity of a test is 0.9 for men and 0.8 for women; the specificity of a test is 0.6 for men and 0.7 for women. The prevalence of disease is 0.1 for both men and women. What is the positive predictive value of the test for men, for women, and overall? Assume $P(\text{men}) = 0.5$.

Q3. For a variety of reasons, self-reported disease outcomes are frequently used without verification in epidemiologic research. In a study by Parikh-Patel et al., researchers looked at the relationship between self-reported cancer cases and actual cases. They used the self-reported cancer data from a California Teachers Study and validated the cancer cases by using the California Cancer Registry data. The following table reports their findings for breast cancer:

Cancer Reported	Cancer in Registry	Cancer Not in Registry	Total
Yes	2991	2244	5235
No	112	115849	115961
Total	3103	118093	121196

1. Let A be the event of reporting breast cancer in the California Teachers Study. Find the probability of A in this study.

2. Let B be the event of having breast cancer confirmed in the California Cancer Registry. Find the probability of B in this study.

3. Find $P(A \cap B)$.

4. Find $P(A|B)$.

5. Find $P(B|A)$.

6. Find the sensitivity of using self-reported breast cancer as a predictor of actual breast cancer in the California registry.

7. Find the specificity of using self-reported breast cancer as a predictor of actual breast cancer in the California registry.