MIE Lecture Notes

Probability and Statistics

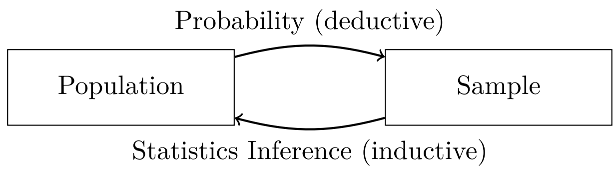
Contents

Chapter 1

Statistics Definitions

Global definition: Statistics involves collecting, organizing, summarizing, presenting, and analyzing data, as well as making inferences, conclusions, and decisions based on data.

Statistical definition: A statistic is a numerical value calculated from data (e.g. mean, proportion, standard deviation).



Basic Terminology

Individuals: Objects on which data are collected (people, animals, plots of land, etc.).
Variable: Any characteristic of an individual.
Population: The entire group of individuals of interest.
Sample: A subset of individuals taken from the population.
Statistical Inference: Drawing conclusions about a population based on a sample.

Sampling Methods

Simple Random Sample (SRS):

- Every possible group of size n has an equal chance of being selected.
- Helps avoid bias in sampling.
- Can be selected using random number tables or software.

Stratified Random Sampling:

- The population is divided into homogeneous groups *(individuals are similar with respect to the variable being studied)* called strata.
- A simple random sample is taken from each stratum. *(one subgroup of the population created)*
- Ensures that important subgroups are neither over nor under represented.

Types of Variables

Categorical Variable: Places individuals into categories (e.g. gender, major). These are qualitative.

Quantitative Variable: Takes numerical values for which arithmetic operations are meaningful.

- Discrete
- Continuous

Distributions

Distribution: Describes what values a variable takes and how often those values occur. When examining a distribution, look for:

- **Shape**
- **Center**
- **Spread**
- **Outliers**

Outlier: An individual value that falls outside the overall pattern of the data.

Describing Distributions with Numbers

Central Tendency: Describes where the data cluster or center.

Central Tendency: Describes where the data cluster or center.

- Mean: average value
- Median: middle value

Mean (Arithmetic Mean):

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Median:

$$\boxed{\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{if } n \text{ is odd} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{if } n \text{ is even} \end{cases}}$$

Theorem 1.1

1. The mean is more sensitive to extreme values than the median.
2. Changing a single data value will always change the mean, but may not change the median.
3. If a distribution is exactly symmetric, the mean and median are equal.

Trimmed Mean: The mean computed after removing extreme values.

$$\boxed{\bar{x}_{\text{trim}} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_{(i)}}$$

where k values are removed from both ends of the ordered data. (normally given in question like 10%)

Measures of Spread

Range: Maximum minus minimum. Very sensitive to extreme values.

Sample Variance: Measures the average squared deviation from the mean.

$$\boxed{s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Standard Deviation: The square root of the sample variance.

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

Degrees of Freedom: The number of independent pieces of information available to estimate variability. For sample variance: $df = n - 1$.

Plots

Scatter Plot: Used to display the relationship between two quantitative variables (x, y). A scatter plot helps identify trends, patterns, and associations between variables.

Stem-and-Leaf Plot: An intermediate step between raw data and a frequency table. Preserves the original data values while showing the distribution.

Stem	Leaf
1	2 4 7
2	1 3 5 8
3	0 4 6

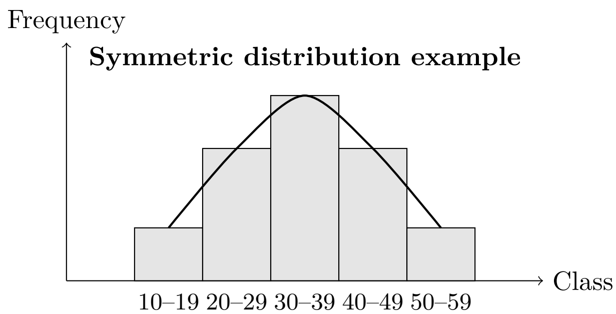
Relative Frequency Table: Shows the proportion of observations in each class.

Class Interval	Class Midpoint	Frequency	Relative Frequency
10–19	14.5	3	0.30
20–29	24.5	4	0.40
30–39	34.5	3	0.30

Histogram: A graphical representation of a frequency or relative frequency table using contiguous bars.

When describing the shape of a histogram, we commonly classify it as:

- **Symmetric**
- **Skewed right** (positively skewed)
- **Skewed left** (negatively skewed)



Chapter 2, Jan 9th

Experiment: A process that generates an outcome.

Sample Space (S): The set of all possible outcomes of an experiment.

Example 1:

Select 3 items from a production line. Each item can be classified as either defective (D) or non-defective (N).

$$S = \{DDD, DDN, DND, NDD, DNN, NDN, NND, NNN\}$$

Since each item has 2 possible outcomes,

$$|S| = 2^3 = 8$$

Example 2:

$$S = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event (A): A subset of the sample space S.

Examples of events:

$$A = \{DDD, DDN, DND, NDD\}$$

$$B = \{NNN\}$$

$$C = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event Operations:

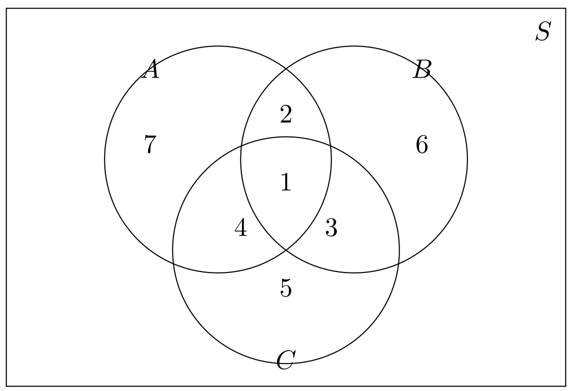
- Complement: A^c (or A')
- Intersection: $A \cap B$
- Union: $A \cup B$
- Null Event: $\varnothing$

If

$$A \cap B = \varnothing,$$

then A and B are mutually exclusive.

Example (Venn Diagram):



$$A = \{DDD, DDN, DND, NDD\}, \quad B = \{NNN\}$$

$$A \cup B = \{DDD, DDN, DND, NDD, NNN\}$$

$$A \cap B = \varnothing$$

Chapter 2: January 12

Review

1. *Experiment: A process that generates an outcome.*
2. *Sample Space (S): The set of all possible outcomes of an experiment.*
3. *Event Operations:*
    - *Complement:* A' (A^c)
    - *Intersection:* $A \cap B$

- *Union:* $A \cup B$
- *Null Event:* $\varnothing$

4. *If $A \cap B = \varnothing$, then A and B are called mutually exclusive.*

$$(A \cap B)' = A' \cup B'$$

$$(A \cup B)' = A' \cap B'$$

$$A \cap \varnothing = \varnothing$$

$$A \cup \varnothing = A$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Probability

$P(A)$ = probability of event A: the proportion of times the event occurs in infinitely many repetitions of the experiment.

Theorem2.1:

$$0 \leq P(A) \leq 1$$

$$P(A) + P(A') = 1$$

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

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$

Mutually Exclusive Events

Definition: If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

If

$$A_1 \cup A_2 \cup \cdots \cup A_n = S,$$

then $\{A_1, A_2, \ldots, A_n\}$ is a partition of S.

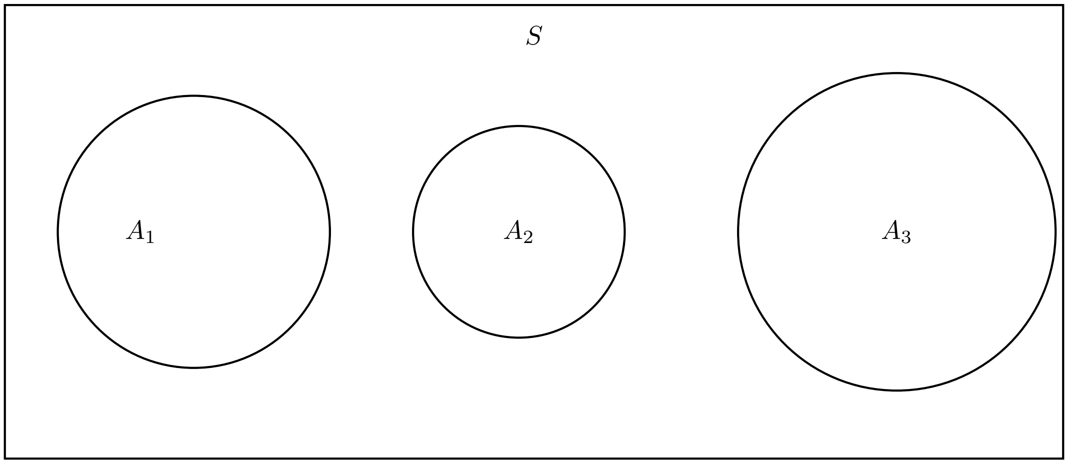


Figure 1: Partition of the sample space S into A_1, A_2, A_3

Example

In a class of 33 students:

- 17 earned an A on the midterm
- 14 earned an A on the final
- 11 earned no A on either exam

Find the probability that a randomly selected student earned A's on <u>both</u> exams.

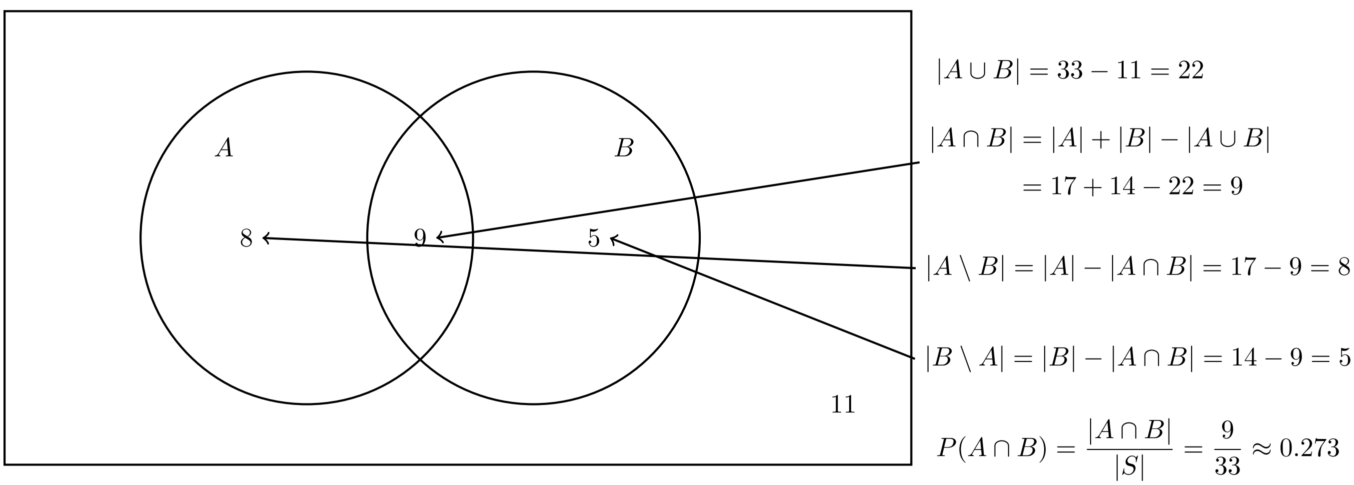


Figure 2: Events A: A on midterm, B: A on final, with region counts and calculations

Theorem 2.2 (Equally Likely Outcomes):

If the sample space S has a finite number of outcomes and all outcomes are equally likely, then for any event A,

$$P(A) = \frac{|A|}{|S|}$$

where
A: the event of interest (a subset of the sample space S),
S: the sample space, i.e. the set of all possible outcomes.

Example 1: Poker Hands Basics

A standard deck has:

$$4 \text{ suits} \times 13 \text{ denominations (A,2,3,\ldots,Q,K)} = 52 \text{ cards.}$$

A poker hand consists of 5 cards chosen from 52:

$$|S| = \binom{52}{5} = 2{,}598{,}960.$$

Combinations Reminder

If there are 3 objects $\{A, B, C\}$ and we choose 2:

$$\binom{3}{2} = \frac{3!}{(3-2)!2!}.$$

Order does not matter.

Example 2: Probability of 2 Aces and 1 Jack

A 5-card hand contains:

- exactly 2 aces,
- exactly 1 jack,
- 2 cards that are neither aces nor jacks.

$$P(2 \text{ aces and } 1 \text{ jack}) = \frac{\binom{4}{2}\binom{4}{1}\binom{44}{2}}{\binom{52}{5}}.$$

Example 3: Probability of a Full House

A full house consists of:

- 3 cards of one denomination
- 2 cards of a different denomination

Number of full house hands:

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}.$$

Thus,

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}.$$

Example 4: Probability of Four of a Kind

A four of a kind consists of:

- 4 cards of the same denomination
- 1 remaining card of a different denomination

Number of such hands:

$$\binom{13}{1}\binom{4}{4}\binom{48}{1}.$$

Thus,

$$P(\text{four of a kind}) = \frac{\binom{13}{1}\binom{4}{4}\binom{48}{1}}{\binom{52}{5}}.$$

Example 5: Probability of Exactly One Pair

An **excatly** one-pair hand consists of:

- 1 pair
- 3 cards of different denominations, none matching the pair

Number of such hands:

$$\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3.$$

Thus,

$$P(\text{exactly one pair}) = \frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3}{\binom{52}{5}}.$$

Note*: Counting Patterns

$$\binom{a}{b}$$

Meaning: Choose b different items from a **at once**, order does not matter.

Key features:

- No repeats
- Grouped choice
- Used when items must be <u>distinct</u>

$$\binom{a}{1}^b$$

Meaning: Make b independent choices, each time choosing 1 item from c.

Key features:

- Repeats allowed
- Choices are independent
- Used when selections do not restrict each other

Rule to Remember:

$$\text{Different items, no repeats} \;\Rightarrow \binom{a}{b}$$

$$\text{Independent choices} \;\Rightarrow \binom{c}{1}^b$$

Chapter 2 continue, Jan 14

Review

1. *Probability is the proportion of times the event occurs in infinitely many repetitions of the experiment.*
2. $0 \le P(A) \le 1$
3. $P(A) + P(A^c) = 1$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. $$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$
6. *Permutation: A permutation counts ordered arrangements.*

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

Example 1: Two fair dice

A pair of fair dice are rolled. Find the probability that the second die lands on a smaller value than the first.

The outcomes where the second die is smaller than the first are represented below.

First Die (Stem)	Second Die (Leaf)
2	1
3	1 2
4	1 2 3
5	1 2 3 4
6	1 2 3 4 5

There are 15 favorable outcomes and 36 total outcomes.

$$P(\text{second} < \text{first}) = \frac{15}{36} = \frac{5}{12}.$$

Conditional Probability

The conditional probability of an event B given that event A has occurred is the probability that B occurs when it is known that A has occurred.

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

Example 2: Drinking Survey

A survey records the following data:

	D	N	Total
M	19	41	60
F	12	28	40
Total	31	69	100

The symbols used above are defined as follows:

- M: male
- F: female
- D: the individual drinks
- N: the individual does not drink

$$P(D|M) = \frac{19}{60} \qquad P(M|D) = \frac{19}{31}$$

Law of Total Probability

Theorem 2.3: If $B_1, \ldots, B_k$ form a partition (do not overlap but covers the whole sample space) of S with $P(B_i) > 0$, then for any event A,

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Example 3: Monty Hall (3 doors)

Car location	Monty opens	Probability	Stay	Switch
Door 1	Door 2	$\frac{1}{6}$	Car	Goat
Door 1	Door 3	$\frac{1}{6}$	Car	Goat
Door 2	Door 3	$\frac{1}{3}$	Goat	Car
Door 3	Door 2	$\frac{1}{3}$	Goat	Car

Staying wins only when the car is behind Door 1, so

$$P(\text{win by staying}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Switching wins when the car is behind Door 2 or Door 3, so

$$P(\text{win by switching}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

Example 4: Birthday Problem

Assume the following:

- Leap years are ignored
- All 365 birthdays are equally likely
- Birthdays of different people are independent

Question: What is the probability that at least two people share the same birthday in a group of n people?

Rather than computing this directly, we use the complement rule.

$$P(\text{at least one match}) = 1 - P(\text{no match})$$

Probability of no shared birthdays

- Person 1 can have any birthday: probability 1
- Person 2 must avoid that birthday: $\frac{364}{365}$
- Person 3 must avoid the first two birthdays: $\frac{363}{365}$
- ...
- Person n must avoid the previous $n-1$ birthdays: $\frac{365-(n-1)}{365}$

Therefore,

$$P(\text{no match}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-(n-1)}{365}$$

or equivalently,

$$P(\text{no match}) = \prod_{k=0}^{n-1} \frac{365-k}{365}$$

Final result

$$P(\text{at least one shared birthday}) = 1 - \prod_{k=0}^{n-1} \frac{365-k}{365}$$

Important values

- For $n = 23$: $\quad P(\text{at least one match}) \approx 0.507$
- For $n = 57$: $\quad P(\text{at least one match}) \approx 0.99$

Chapter 2 — Jan 16

Review: Conditional Probability

Conditional Probability:

The probability of event B given that event A has occurred is

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

Read as: the probability of B given A.

Independence of Events

Definition (Independence): Events A and B are independent if and only if

$$\boxed{P(B \mid A) = P(B)}$$

Equivalently,

$$P(A \mid B) = P(A)$$

or

$$\boxed{P(A \cap B) = P(A)\,P(B)}$$

Multiple Independent Events

Definition:

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

$$\boxed{P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)\,P(A_2) \cdots P(A_k)}$$

Mutual Independence

Mutual Independence : A collection of events $A_1, A_2, \ldots, A_n$ is mutually independent if and only if for *every* subcollection $\{A_{i_1}, \ldots, A_{i_k}\}$,

$$\boxed{P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})}$$

Example (Three Events):

Events A_1, A_2, A_3 are mutually independent if all of the following hold:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

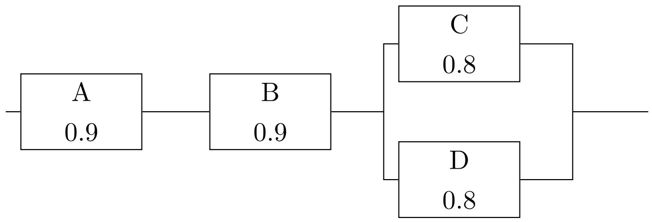
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

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

Note: Mutually exclusive events are dependent. If one event occurs, the other cannot occur.

Example: Component Reliability

An electrical system has four components A, B, C, D. The system works if A and B work and at least one of C or D works. Assume all components are independent.



$$P(A) = 0.9, \quad P(B) = 0.9, \quad P(C) = 0.8, \quad P(D) = 0.8$$

(a) Probability the entire system works

The system works if A and B work and either C or D works.

$$\begin{aligned} P(\text{system works}) &= P(\text{all work}) + P(A, B, C \text{ work}, D \text{ does not}) \\ &\quad + P(A, B, D \text{ work}, C \text{ does not}) \end{aligned}$$

$$= (0.9)(0.9)(0.8)(0.8) + (0.9)(0.9)(0.8)(1 - 0.8) + (0.9)(0.9)(0.8)(1 - 0.8)$$

$$\boxed{P(\text{system works}) = 0.7776}$$

(b) Conditional probability

$$P(C^c \mid \text{system works}) = \frac{P(C^c \cap \text{system works})}{P(\text{system works})}$$

$$P(C^c \cap \text{system works}) = (0.9)(0.9)(0.8)(1 - 0.8)$$

$$\boxed{P(C^c \mid \text{system works}) = \frac{(0.9)(0.9)(0.8)(1 - 0.8)}{0.7776} = 0.16}$$

Theorem of Total Probability

Let $B_1, B_2, \ldots, B_k$ be a partition of the sample space S with $P(B_i) > 0$ for all i. Then for any event $A \subseteq S$,

$$\boxed{P(A) = \sum_{i=1}^{k} P(A \mid B_i)\, P(B_i) = \sum_{i=1}^{k} P(A \cap B_i)}$$

Theorem: Bayes' Rule (1701–1761)

Let $B_1, B_2, \ldots, B_k$ be a partition of the sample space S such that $P(B_i) > 0$ for $i = 1, \ldots, k$. For any event $A \subseteq S$ (in) with $P(A) > 0$,

$$\boxed{P(B_r \mid A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)\,P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i)\,P(A \mid B_i)}, \qquad r = 1, \ldots, k}$$

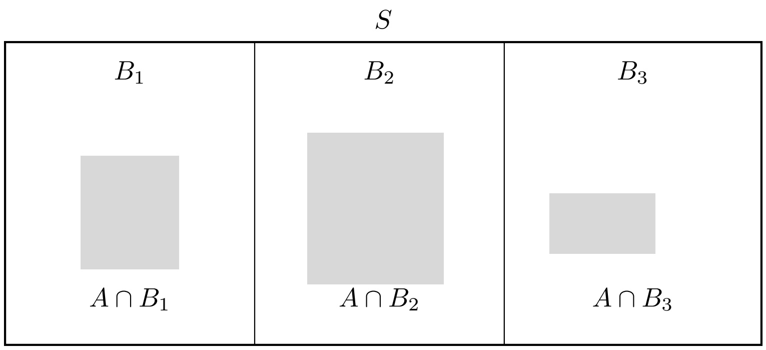


Figure 3: Visual interpretation of Bayes' Rule

Example (Medical Test)

The fraction of people in a population who have a certain disease is 0.01.

$$P(D) = 0.01, \quad P(D^c) = 0.99$$

The test characteristics are:

$$P(\text{test says } D \mid D^c) = 0.05 \quad \text{(false positive rate)}$$

$$P(\text{test says } D^c \mid D) = 0.20 \quad \text{(false negative rate)}$$

Thus,

$$P(\text{test says } D \mid D) = 1 - 0.20 = 0.80$$

Note: $1 - P(\textit{test says } D^c \mid D)$ *is called the* *sensitivity* *of the test, and* $1 - P(\textit{test says } D \mid D^c)$ *is called the* *specificity*.

(a) Probability the test says disease

$$P(\text{test says } D) = P(D \cap \text{test says } D) + P(D^c \cap \text{test says } D)$$

$$= P(\text{test says } D \mid D)P(D) + P(\text{test says } D \mid D^c)P(D^c)$$

$$= (0.80)(0.01) + (0.05)(0.99) = 0.0575$$

(b) Probability of disease given positive test

$$P(D \mid \text{test says } D) = \frac{P(D \cap \text{test says } D)}{P(\text{test says } D)}$$

$$= \frac{P(\text{test says } D \mid D)P(D)}{0.0575} = \frac{(0.80)(0.01)}{0.0575}$$

$$\boxed{P(D \mid \text{test says } D) \approx 0.139}$$

(c) Probability of disease given negative test

$$P(D \mid \text{test says } D^c) = \frac{P(D \cap \text{test says } D^c)}{P(\text{test says } D^c)}$$

$$= \frac{P(\text{test says } D^c \mid D)P(D)}{1 - P(\text{test says } D)}$$

$$= \frac{(0.20)(0.01)}{1 - 0.0575}$$

$$\boxed{P(D \mid \text{test says } D^c) \approx 0.00212}$$