Lecture 1:
Statistics and Probability

Why learn statistics and probability?

It helps you answer the following questions after the course:

- If you conduct a COVID-19 test, your result is POSITIVE. According to the accuracy of the test, do you know the chance that you indeed get COVID-19? (Bayes' Theorem)
- If you are working on the survey department of a company and you are asked to estimate the average weight of an adult male in Hong Kong using the simple random sampling method, how can you express the precision/error and certainty/uncertainty associated with your sample average weight? (Estimation in Statistics)
- If you are an engineer making a soft-drink vending machine, how do you know whether the machine is correctly calibrated, i.e., producing the right amount of soft drink? (Procedure for a Statistical Test)

Why use online learning materials?

- They are easy to access.
- They are free to use in this course and even after your graduation.
- Online supports (e.g., calculators, experiments) are fast, accurate, and user-friendly.
- All the learning information is safely stored in an online database.
- The online materials keep continuously updated.
- You can continue to use them to learn the subject after this course.

Outline

- Introduction to statistics and probability and relationship between them
- Introduction to basic concepts of data statistics: variable, population, sample, mean, variance, standard deviation, simple random sampling, law of large numbers, statistical experiment
- Introduction to basic concepts of probability: sets, subsets, element, sample point, event, sample space, set operations (union, intersection, complement), probability of an event
- Rules of Probability: subtraction, multiplication, addition, probability of a sample point, Bayes' theorem

Statistics

- **Statistics** is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data.
- In applying statistics to a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model to be studied.
- Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.

https://en.wikipedia.org/wiki/Statistics

Probability

- **Probability** is the branch of mathematics concerning numerical descriptions of how likely an event is to occur or how likely it is that a proposition is true.
- The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty.
- The higher the probability of an event, the more likely it is that the event will occur.
- A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since no other outcomes are possible, the probability of either "heads" or "tails" is 1/2 (which could also be written as 0.5 or 50%).

https://en.wikipedia.org/wiki/Probability

Relationship between Statistics and Probability

- In statistics, we apply probability (probability theory) to draw conclusions from data.
- **Statistics example:** You have a coin of an unknown source. To investigate whether it is fair you toss it 100 times and count the number of heads. Let's say you count 60 heads. Your job as a statistician is to draw a conclusion (inference) from this data.
- **Probability example:** You have a fair coin (equal probability of heads or tails). You will toss it 100 times. What is the probability of 60 or more heads? We can get only a single answer because of the standard computation strategy.

https://towardsdatascience.com/probability-vs-statistics-95f221cc74f7

The Basic for Exploring Data

What Are Variables?

In statistics, a **variable** has two defining characteristics:

- A variable has an attribute that describes a person, place, thing, or idea.
- The value of the variable can "vary" from one entity to another.

For example, a person's *hair color* is a potential variable, which could have the value of "black" for one person and "brown" for another.

Source: https://stattrek.com/

Population vs Sample

The main difference between a population and sample has to do with how observations are assigned to the data set.

- A **population** includes all of the elements from a set of data.
- A **sample** consists one or more observations drawn from the population.

Depending on the sampling method, a sample can have fewer observations than the population, the same number of observations, or more observations.

Source: https://stattrek.com/

Population vs Sample (continued)

Other differences have to do with terminology, notation, and computations. For example,

- A measurable characteristic of a **population**, such as a mean or standard deviation, is called a **parameter**; but a measurable characteristic of a **sample** is called a **statistic**.
- We will see in future lessons that the mean of a population is denoted by the symbol μ; but the mean of a sample is denoted by the symbol $\bar{x}$.
- We will also learn later in this lesson that the formula for the standard deviation of a population is different from the formula for the standard deviation of a sample.

Source: https://stattrek.com/

What is Simple Random Sampling?

A **sampling method** is a procedure for selecting sample elements from a population. **Simple random sampling** refers to a sampling method that has the following properties.

- The population consists of *N* objects.
- The sample consists of *n* objects.
- All possible samples of *n* objects are equally likely to occur.

An important benefit of simple random sampling is that it allows researchers to use statistical methods to analyze sample results. For example, given a simple random sample, researchers can use statistical methods to define a confidence interval around a sample mean. Statistical analysis is not appropriate when non-random sampling methods are used.

There are many ways to obtain a simple random sample. One way would be the lottery method. Each of the *N* population members is assigned a unique number. The numbers are placed in a bowl and thoroughly mixed. Then, a blind-folded researcher selects *n* numbers. Population members having the selected numbers are included in the sample.

Source: https://stattrek.com/

The Mean

The **mean** of a sample or a population is computed by adding all of the observations and dividing by the number of observations. If we have five teenage boys with weight 100, 100, 130, 140 and 150 (pounds), the mean weight would equal (100 + 100 + 130 + 140 + 150)/5 = 620/5 = 124 pounds. In the general case, the mean can be calculated, using one of the following equations:

$$\text{Population mean} = \mu = \Sigma X / N$$

$$\text{Sample mean} = \bar{x} = \Sigma x / n$$

where ΣX is the sum of all the population observations, N is the number of population observations, Σx is the sum of all the sample observations, and n is the number of sample observations.

When statisticians talk about the mean of a population, they use the Greek letter μ to refer to the mean score. When they talk about the mean of a sample, statisticians use the symbol $\bar{x}$ to refer to the mean score.

Source: https://stattrek.com/

The Variance

In a population, **variance** is the average squared deviation from the population mean, as defined by the following formula:

$$\sigma^2 = \Sigma (X_i - \mu)^2 / N$$

where σ^2 is the **population variance**, μ is the population mean, X_i is the *i* th element from the population, and *N* is the number of elements in the population.

Observations from a simple random sample can be used to estimate the variance of a population. For this purpose, sample variance is defined by slightly different formula, and uses a slightly different notation:

$$s^2 = \Sigma (x_i - \bar{x})^2 / (n - 1)$$

where s^2 is the **sample variance**, $\bar{x}$ is the sample mean, x_i is the *i* th element from the sample, and *n* is the number of elements in the sample.

Source: https://stattrek.com/

The Standard Deviation

The **standard deviation** is the square root of the variance. Thus, the standard deviation of a population is:

$$\sigma = \text{sqrt}\left[\sigma^2 \right] = \text{sqrt}\left[\Sigma \left(X_i - \mu \right)^2 / N \right]$$

where σ is the **population standard deviation**, μ is the population mean, X_i is the *i*th element from the population, and N is the number of elements in the population.

Statisticians often use simple random samples to estimate the standard deviation of a population, based on sample data. Given a simple random sample, the best estimate of the standard deviation of a population is:

$$s = \text{sqrt}\left[s^2 \right] = \text{sqrt}\left[\Sigma \left(x_i - \bar{x} \right)^2 / \left(n - 1 \right) \right]$$

where s is the **sample standard deviation**, $\bar{x}$ is the sample mean, x_i is the *i*th element from the sample, and n is the number of elements in the sample.

Source: https://stattrek.com/

Probability

What is Probability?

The **probability** of an event refers to the likelihood that the event will occur.

Applied researchers make decisions under uncertainty. Probability theory makes it possible for researchers to quantify the extent of uncertainty inherent in their conclusions and inferences.

Source: https://stattrek.com

How to Interpret Probability

Mathematically, the probability that an event will occur is expressed as a number between 0 and 1. Notationally, the probability of event A is represented by P(A).

- If P(A) equals zero, event A will definitely not occur.
- If P(A) is close to zero, there is only a small chance that event A will occur.
- If P(A) equals 0.5, there is a 50-50 chance that event A will occur.
- If P(A) is close to one, there is a strong chance that event A will occur.
- If P(A) equals one, event A will definitely occur.

In a statistical experiment, the sum of probabilities for all possible outcomes is equal to one. This means, for example, that if an experiment can have three possible outcomes (A, B, and C), then P(A) + P(B) + P(C) = 1.

Source: https://stattrek.com

How to Compute Probability

- Sometimes, a statistical experiment can have *n* possible outcomes, each of which is equally likely. Suppose a subset of *r* outcomes are classified as "successful" outcomes.
- The probability that the experiment results in a successful outcome (S) is:
- P(S) = (Number of successful outcomes) / (Total number of equally likely outcomes) = *r* / *n*
- Consider the following experiment. An urn has 10 marbles. Two marbles are red, three are green, and five are blue. If an experimenter randomly selects 1 marble from the urn, what is the probability that it will be green?
- In this experiment, there are 10 equally likely outcomes, three of which are green marbles. Therefore, the probability of choosing a green marble is 3/10 or 0.30.

Source: https://stattrek.com

How to Compute Probability: Law of Large Numbers

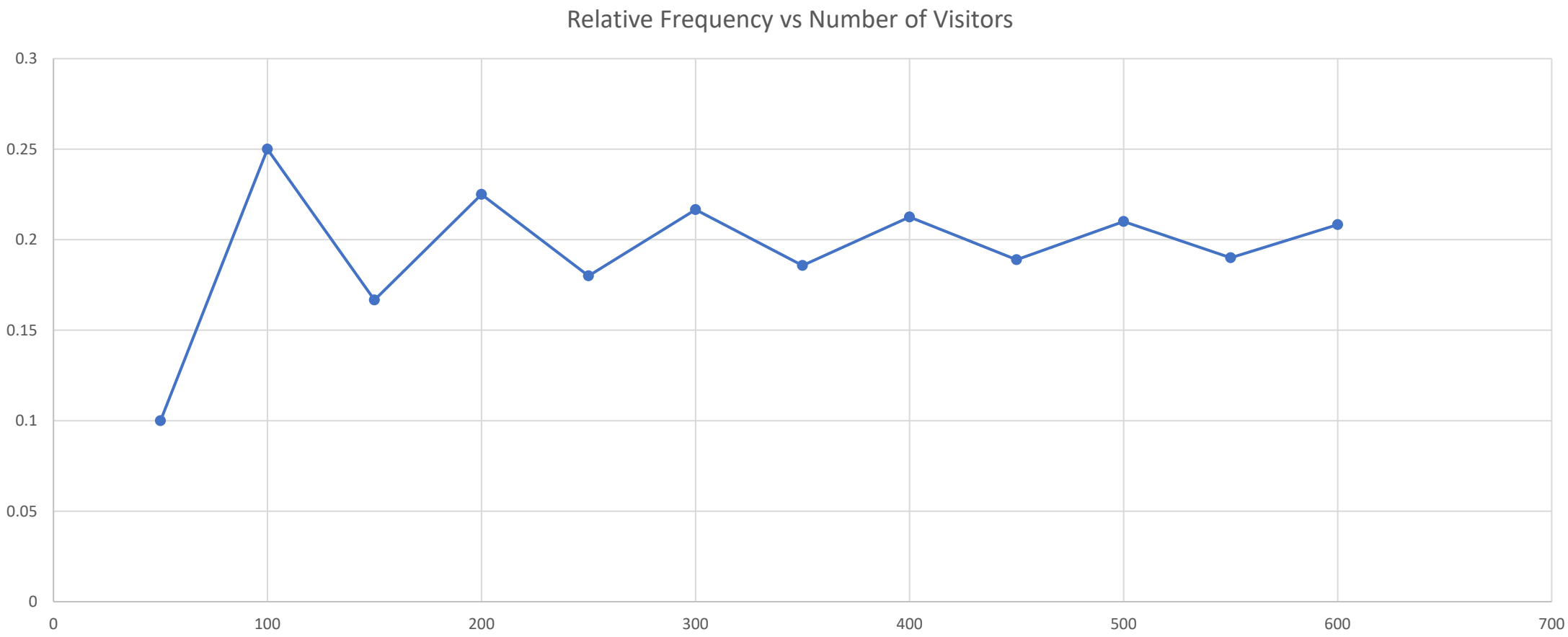
- One can also think about the probability of an event in terms of its *long-run* relative frequency. The relative frequency of an event is the number of times an event occurs, divided by the total number of trials.

$$P(A) = (\text{Number of Occurrence of Event A}) / (\text{Number of Trials})$$

- For example, a merchant notices one day that 5 out of 50 visitors to her store make a purchase. The next day, 20 out of 50 visitors make a purchase. The two relative frequencies (5/50 or 0.10 and 20/50 or 0.40) differ. However, summing results over many visitors, she might find that the probability that a visitor makes a purchase gets closer and closer to 0.20.

Source: https://stattrek.com

How to Compute Probability: Law of Large Numbers (continued)



Source: https://stattrek.com

How to Compute Probability: Law of Large Numbers (continued)

- The previous scatterplot shows the relative frequency of purchase as the number of trials (in this case, the number of visitors) increases. Over many trials, the relative frequency converges toward a stable value (0.20), which can be interpreted as the probability that a visitor to the store will make a purchase.
- The idea that the relative frequency of an event will converge on the probability of the event, as the number of trials increases, is called the law of large numbers.

Source: https://stattrek.com

Test Your Understanding (Polling 1)

Problem

A coin is tossed three times. What is the probability that it lands on heads *exactly* one time?

(A) 0.125
(B) 0.250
(C) 0.333
(D) 0.375
(E) 0.500

Source: https://stattrek.com

Sets and Subsets

Set Definitions

- A **set** is a well-defined collection of objects.
- Each object in a set is called an **element** of the set.
- Two sets are **equal** if they have exactly the same elements in them.
- A set that contains no elements is called a **null set** or an **empty** set.
- If every element in Set *A* is also in Set *B*, then Set *A* is a **subset** of Set *B*.

Source: https://stattrek.com

Set Notation

- A set is usually denoted by a capital letter, such as *A*, *B*, or *C*.
- An element of a set is usually denoted by a small letter, such as *x*, *y*, or *z*.
- A set may be described by listing all of its elements enclosed in braces. For example, if Set *A* consists of the numbers 2, 4, 6, and 8, we may say: *A* = {2, 4, 6, 8}.
- The null set is denoted by {} or ∅.
- Sets may also be described by stating a rule. We could describe Set *A* from the previous example by stating: Set *A* consists of all the even single-digit positive integers.

Source: https://stattrek.com

Sets and Probability

- As we learned previously, probability is all about statistical experiments. When a researcher conducts a statistical experiment, he or she cannot know the outcome in advance. The outcome is determined by chance.
- However, if the researcher can list all the possible outcomes of the experiment, it may be possible to compute the probability of a particular outcome. The list of all possible outcomes from a statistical experiment is called the **sample space**. And a particular outcome or collection of outcomes is called an **event**.
- You can see that a sample space is a type of set. It is a well-defined listing of all possible outcomes from a statistical experiment. And an event in a statistical experiment is a subset of the sample space.

Source: https://stattrek.com

Set Operations

Suppose we have a sample space S defined as follows: S = {1, 2, 3, 4, 5, 6}. Within that sample space, suppose we define two subsets as follows: X = {1, 2} and Y= {2, 3, 4}.

- The **union** of two sets is the set of elements that belong to one or both of the two sets. Thus, if X is {1, 2} and Y is {2, 3, 4}, the union of sets X and Y is:

$$X \cup Y = \{1, 2, 3, 4\}$$

  Symbolically, the union of X and Y is denoted by $X \cup Y$.

- The **intersection** of two sets is the set of elements that are common to both sets. Thus, if X is {1, 2} and Y is {2, 3, 4}, the intersection of sets X and Y is:

$$X \cap Y = \{2\}$$

  Symbolically, the intersection of X and Y is denoted by $X \cap Y$.

Source: https://stattrek.com

Set Operations (continued)

- The **complement** of an event is the set of all elements in the sample space but not in the event. Thus, if the sample space is {1, 2, 3, 4, 5, 6}, and Y is {2, 3, 4}, the complement of set Y is:

$$Y' = \{1, 5, 6\}$$

Here, we denote the complement of set Y as Y'. In other places, you may see the complement of set Y denoted as Y^c.

Source: https://stattrek.com

Sample Problems

1. Describe the set of vowels.

2. Describe the set of positive integers.

3. Set A = {1, 2, 3} and Set B = {3, 2, 1}. Is Set A equal to Set B?

4. What is the set of men with four arms?

5. Set A = {1, 2, 3} and Set B = {1, 2, 4, 5, 6}. Is Set A a subset of Set B?

Source: https://stattrek.com

What is a Statistical Experiment?

All **statistical experiments** have three things in common:

- The experiment can have more than one possible outcome.
- Each possible outcome can be specified in advance.
- The outcome of the experiment depends on chance.

Example: A coin toss has all the attributes of a statistical experiment. There is more than one possible outcome. We can specify each possible outcome (i.e., heads or tails) in advance. And there is an element of chance, since the outcome is uncertain.

Source: https://stattrek.com

The Sample Space

- A **sample space** is a set of elements that represents all possible outcomes of a statistical experiment.
- A **sample point** is an element of a sample space.
- An **event** is a subset of a sample space - one or more sample points.

Source: https://stattrek.com

Probability of an Event

With some statistical experiments, each sample point is equally likely to occur. In this situation, the probability of an event is very easy to compute. It is:

$$P(E) = \frac{\text{Number of sample points in event}}{\text{Number of sample points in sample space}}$$

Think about the toss of a single die. The sample space consists of six possible outcomes (1, 2, 3, 4, 5, and 6). And each outcome is equally likely to occur. Suppose we defined Event A to be the die landing on an odd number. There are three odd numbers (1, 3, and 5). So, the probability of Event A would be 3/6 or 0.5.

Source: https://stattrek.com

Types of events

- Two events are **mutually exclusive** if they have no sample points in common.
- Two events are **independent** when the occurrence of one does not affect the probability of the occurrence of the other.

Source: https://stattrek.com

Test Your Understanding

1. Suppose I roll a die. Is that a statistical experiment?
2. When you roll a single die, what is the sample space?
3. Which of the following are sample points when you roll a die - 3, 6, and 9?
4. Which of the following sets represent an event when you roll a die?
   A. {1}
   B. {2, 4,}
   C. {2, 4, 6}
   D. All of the above
5. Consider the events listed below. Which are mutually exclusive?
   A. {1}
   B. {2, 4,}
   C. {2, 4, 6}
6. Suppose you roll a die two times. Is each roll of the die an independent event?

Source: https://stattrek.com

Rules of Probability

Often, we want to compute the probability of an event from the known probabilities of other events. This lesson covers some important rules that simplify those computations.

Definitions and Notation

- Two events are **mutually exclusive** or **disjoint** if they cannot occur at the same time.
- The probability that Event A occurs, given that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).
- The **complement** of an event is the event not occurring. The probability that Event A will not occur is denoted by P(A').

Source: https://stattrek.com

Definitions and Notation (continued)

- The probability that Events A **and** B ***both*** occur is the probability of the **intersection** of A and B. The probability of the intersection of Events A and B is denoted by $P(A \cap B)$. If Events A and B are mutually exclusive, $P(A \cap B) = 0$.
- The probability that Events A **or** B occur is the probability of the **union** of A and B. The probability of the union of Events A and B is denoted by $P(A \cup B)$.
- If the occurrence of Event A changes the probability of Event B, then Events A and B are **dependent**. On the other hand, if the occurrence of Event A does not change the probability of Event B, then Events A and B are **independent**.

Source: https://stattrek.com

Rule of Subtraction

Previously, we learned two important properties of probability:

- The probability of an event ranges from 0 to 1.
- The sum of probabilities of all possible events equals 1.

The rule of subtraction follows directly from these properties.

Rule of Subtraction. The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

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

Example: Suppose, for example, the probability that Bill will graduate from college is 0.80. What is the probability that Bill will not graduate from college? Based on the rule of subtraction, the probability that Bill will not graduate is 1.00 - 0.80 or 0.20.

Source: https://stattrek.com

Rule of Multiplication

- The rule of multiplication applies to the situation when we want to know the probability of the intersection of two events; that is, we want to know the probability that two events (Event A and Event B) both occur.

- **Rule of Multiplication** The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.
    - $P(A \cap B) = P(A)\ P(B|A)$

Source: https://stattrek.com

Rule of Multiplication (continued)

Example

An urn contains 6 red marbles and 4 blue marbles. Two marbles are drawn *without replacement* from the urn. What is the probability that both of the marbles are blue?

Solution: Let A = the event that the first marble is blue; and let B = the event that the second marble is blue. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are blue. Therefore, P(A) = 4/10.
- After the first selection, there are 9 marbles in the urn, 3 of which are blue. Therefore, P(B|A) = 3/9.

Therefore, based on the rule of multiplication:

$$P(A \cap B) = P(A)\, P(B|A)$$
$$P(A \cap B) = (4/10) * (3/9) = 12/90 = 2/15 = 0.133$$

Source: https://stattrek.com

Rule of Addition

The rule of addition applies to the following situation. We have two events, and we want to know the probability that either event occurs.

Rule of Addition The probability that Event A **or** Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

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

Note: Invoking the fact that $P(A \cap B) = P(A)P(B \mid A)$, the Addition Rule can also be expressed as:

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

Source: https://stattrek.com

Rule of Addition (continued)

Example

A student goes to the library. The probability that she checks out (a) a work of fiction is 0.40, (b) a work of non-fiction is 0.30, and (c) both fiction and non-fiction is 0.20. What is the probability that the student checks out a work of fiction, non-fiction, or both? Note that we assume that a work can be classified as both fiction and non-fiction.

Solution: Let F = the event that the student checks out fiction; and let N = the event that the student checks out non-fiction. Then, based on the rule of addition:

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$
$$P(F \cup N) = 0.40 + 0.30 - 0.20 = 0.50$$

Source: https://stattrek.com

Test Your Understanding (Polling 2)

Problem 1

An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *with replacement* from the urn. What is the probability that both of the marbles are black?

(A) 0.16
(B) 0.32
(C) 0.36
(D) 0.40
(E) 0.60

Source: https://stattrek.com

Test Your Understanding (Polling 3)

Problem 2

A card is drawn randomly from a deck of ordinary playing cards. You win $10 if the card is a spade **or** an ace. What is the probability that you will win the game?

(A) 1/13
(B) 13/52
(C) 4/13
(D) 17/52
(E) None of the above.

Source: https://stattrek.com

How to Solve Probability Problems

You can solve many simple probability problems just by knowing two simple rules:

- The probability of any sample point can range from 0 to 1.
- The sum of probabilities of all sample points in a sample space is equal to 1.

The following sample problems show how to apply these rules to find (1) the probability of a sample point and (2) the probability of an event.

Source: https://stattrek.com

Probability of a Sample Point

The **probability** of a sample point is a measure of the likelihood that the sample point will occur.

Example 1
Suppose we conduct a simple statistical experiment. We flip a coin one time. The coin flip can have one of two equally-likely outcomes - heads or tails. Together, these outcomes represent the sample space of our experiment. Individually, each outcome represents a sample point in the sample space. What is the probability of each sample point?

Solution: The sum of probabilities of all the sample points must equal 1. And the probability of getting a head is equal to the probability of getting a tail. Therefore, the probability of each sample point (heads or tails) must be equal to 1/2.

Source: https://stattrek.com

Probability of a Sample Point (continued)

Example 2
Let's repeat the experiment of Example 1, with a die instead of a coin. If we toss a fair die, what is the probability of each sample point?

Solution: For this experiment, the sample space consists of six sample points: {1, 2, 3, 4, 5, 6}. Each sample point has equal probability. And the sum of probabilities of all the sample points must equal 1. Therefore, the probability of each sample point must be equal to 1/6.

Source: https://stattrek.com

Bayes' theorem (aka Bayes' rule)

Bayes' theorem (also known as Bayes' rule) is a useful tool for calculating conditional probabilities. Bayes' theorem can be stated as follows:

Bayes' theorem.. Let $A_1, A_2, \ldots, A_n$ be a set of mutually exclusive events that together form the sample space S. Let B be any event from the same sample space, such that P(B) > 0. Then,

$$P(A_k \mid B) = \frac{P(A_k \cap B)}{[P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)]}$$

Note: Invoking the fact that $P(A_k \cap B) = P(A_k)P(B \mid A_k)$, Baye's theorem can also be expressed as

$$P(A_k \mid B) = \frac{P(A_k) P(B \mid A_k)}{[P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2) + \ldots + P(A_n) P(B \mid A_n)]}$$

Source: https://stattrek.com

Test Your Understanding

Example

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Source: https://stattrek.com

Test Your Understanding

Solution: The sample space is defined by two mutually-exclusive events - it rains, or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.

- Event A_1. It rains on Marie's wedding.
- Event A_2. It does not rain on Marie's wedding.
- Event B. The weatherman predicts rain.

Source: https://stattrek.com

Test Your Understanding (continued)

In terms of probabilities, we know the following:

- $P(A_1) = 5/365 = 0.0136985$ [It rains 5 days out of the year.]
- $P(A_2) = 360/365 = 0.9863014$ [It does not rain 360 days out of the year.]
- $P(B \mid A_1) = 0.9$ [When it rains, the weatherman predicts rain 90% of the time.]
- $P(B \mid A_2) = 0.1$ [When it does not rain, the weatherman predicts rain 10% of the time.]

Source: https://stattrek.com

Test Your Understanding (continued)

We want to know $P(A_1 \mid B)$, the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from Bayes' theorem, as shown below.

$$P(A_1 \mid B) = \frac{P(A_1) P(B \mid A_1)}{P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2)}$$

$$P(A_1 \mid B) = \frac{(0.014)(0.9)}{[(0.014)(0.9) + (0.986)(0.1)]}$$

$$P(A_1 \mid B) = 0.111$$

Note the somewhat unintuitive result. Even when the weatherman predicts rain, it rains only about 11% of the time. Despite the weatherman's gloomy prediction, there is a good chance that Marie will not get rained on at her wedding.

Source: https://stattrek.com

EE1004 Teaching and Learning Survey (14 January 2022)

https://www.questionpro.com/a/SurveyPreview