Date: Jan 11, 2022

Biostatistics?

Biostatistics is the branch of applied statistics with applications in the health sciences and biology.

Why biostatistics?

Many statistical methods are more heavily used in health sciences than elsewhere.
For example, false positive/ false negative, survival analysis, prevalence of disease, measures of risk, etc. are everyday discussed topic in health sciences.

What are we going to learn in this course?

- ***Probability used in measuring risk (Odds ratio, relative risk, survival function, false positive/negative etc. are direct application of probability)***
- ***Measures of risk***
- ***Survival Analysis***
- ***Regression for measuring risk of certain factor in relation to the development of disease etc.***

Today, we will start reviewing probability: Part of Chapter 3 of Text Book.

3.1 What is probability?

Probability is the measure ***chance or likelihood*** of an event.

For example, we might be interested to know:

- what is the probability of getting a Covid-19 from a grocery store?
- What is probability of getting hospitalized with covid-19?
- What is probability of getting Covid-19 given vaccinated?
- Is the event of getting Covid-19 in vaccinated and unvaccinated population independent?

Let us visit **https://www.uofmhealth.org/coronavirus/covid19-numbers**

As of January 10, 2022, total patients hospitalized in Michigan Medicine with Covid-19 is 128, of whom 49 are vaccinated and 79 are unvaccinated.

$$P(Covid-19\ among\ vac) = 49/128 = 0.38$$
$$P(Covid-19\ among\ unvac) = 99/128 = 0.62$$

3.2 How to measure probability?

The computational formula for probability is related with the elements of p**robability.**

What are **elements of probability**?

Elements of probability of an event:

- Experiment (an act whose outcome cannot be known until the experiment is completed)
- sample space (the set of all possible outcome of an experiment)
- events (any collection of outcomes of the sample space)

Example K.1

Suppose that a couple wishes to be tested for Covid-19 given some symptoms. The result of the test for either of the couple may be a Positive (D) or a Negative (N).

Then,

(a) What is the experiment in this example?
(b) What would be the sample space?
(c) How many outcomes will be in the sample space?
(d) Define an event with outcomes referring to negative, for the couple.
(e) Define the probability that none of the couple is positive.
(f) What is the probability the couple will be tested as a positive or negative?
(g) What is the maximum and minimum value of probability of an event?

Remarks:

- Letter S is used to represent a sample space. $n(S)$ refers to the number of outcomes in the sample.
- Any other letter can be used for an event other than the sample space.
- Sample space itself is an event, called ***certain or sure event***.
- In mathematics or statistics, an empty set, which contains no outcomes from the sample space can is called an ***impossible event***.

Solution (Example K.1)

(a) Get a Covid-19 test.

(b) $S = \{DD, DN, ND, NN\}$

(c) $n(S) = 4$.

(d) $A = \{NN\}$ and $n(A) = 1$.

(e) $P(A) = \frac{n(A)}{n(S)} = \frac{1}{4} = 0.25$

(f) $P(S) = \frac{n(S)}{n(S)} = 1$

(g) Maximum probability is 1, associated with a sure event, as in (f).

The minimum probability is 0, associated with an impossible event.

By denoting an impossible event by $\emptyset$, and $n(\emptyset) = 0$, and $P(\emptyset) = 0$

Probability of an event A

The probability of any event A, denoted by $P(A)$, is defined as follows:

$$P(A) = \frac{n(A)}{n(S)} = \frac{number\ of\ outcomes\ favor\ the\ event\ A}{number\ of\ outcomes\ in\ S}$$

Properties of probability:

1) For any event A, $0 \leq P(A) \leq 1$.

2) For any event A, the complement of the event A, denoted by $\bar{A}$, has the probability

$$P(\bar{A}) = 1 - P(A)$$

Example K.2

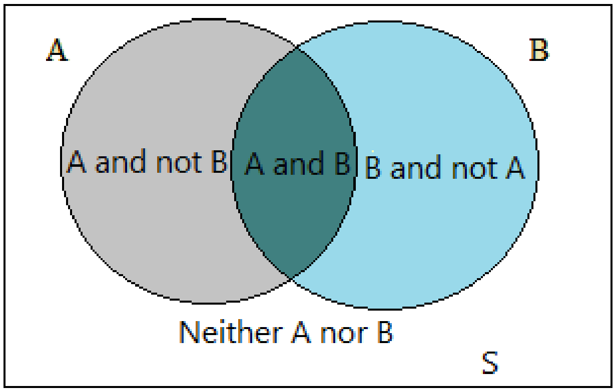
Justify that sure event has probability 1, and impossible event has probability of 0.
Note that $P(S) = \frac{n(S)}{n(S)} = 1$, and $\emptyset = \bar{S}$.
Therefore, $P(\emptyset) = P(\bar{S}) = 1 - P(S) = 1 - 1 = 0$.

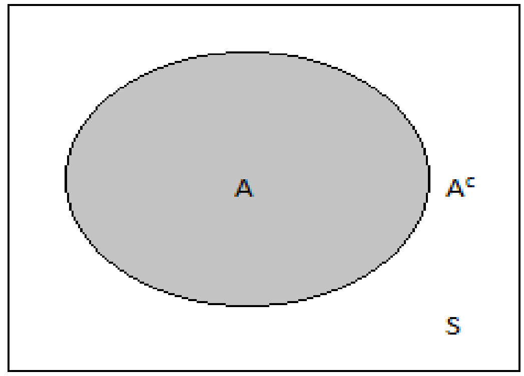
Types of events

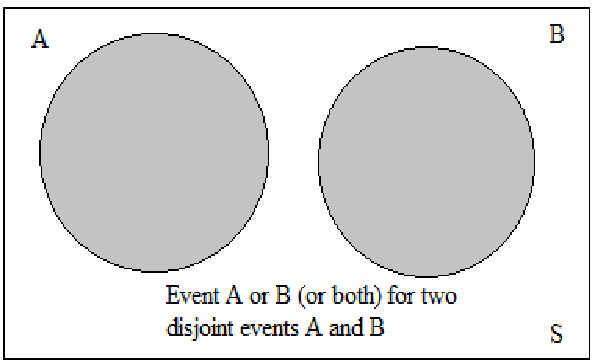
- For any event A, the complement of the event A, denoted by $\bar{A}$, refers to the event $not\ A$, but still part of the sample space.
- Two events A and B are **disjoint** or **mutually exclusive** if they cannot occur together. In this case, the two events A and B have no common outcome, and $A \cap B = \emptyset$. If two events are mutually exclusive, they do not overlap.

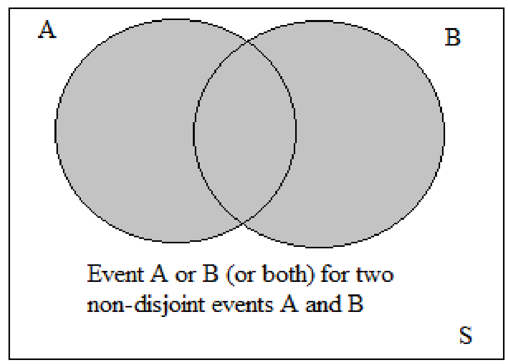
- Two event A and B are non-mutually exclusive or non-disjoint if they have common outcomes between them. In this case, $A \cap B \neq \emptyset$.

  Below we demonstrate various events of common interest using Venn diagrams.









3.3 Useful Notations

Definition 3.4: $\boldsymbol{A \cup B \equiv A \text{ or } B \equiv \text{ at least one of events } A \text{ and } B}$: refers to the event A or B, or both will occur.

Definition 3.5: $\boldsymbol{A \cap B \equiv A \text{ and } B}$: refers to the intersection of events A and B; they occur simultaneously.

Section 3.4: Multiplicative rule and independence

Definition 3.7: If two events A and B are independent if $P(A \cap B) = P(A) \times P(B)$.

Definition 3.8: Two events A and B are dependent if $P(A \cap B) \neq P(A) \times P(B)$.

Example K.3

Among those diagnosed with diabetes in the United States, 26% is treated by insulin, 72% is treated by oral medication and 14% is treated by both insulin and oral medication.

Are the events of those treated by insulin and oral medication independent?

Solution:

Let us define events

I: the person is treated by insulin

O: the person is treated by oral medication

Then, $P(I) = 0.26, P(O) = 0.72$ and $P(I \cap O) = 0.14$

Note that $P(I)P(O) = 0.26 * 0.72 = 0.1872 \neq P(I \cap O)$.

Therefore, the two events are not independent; they are dependent.

However, the event of two persons treated by insulin are intendent, because treatment option of one person is independent of other.

Example K.4

Given the scenarios of previous example, 2 persons with diabetes are randomly selected.

(a) What is the probability that none of them receives insulin?

(b) (ii) What is the probability that both of them receive insulin?

Solution

(a) $0.74 * 0.74 = 0.74^2 = 0.5476$

(b) $0.26^2 = 0.0676$

Equation 3.2 Multiplicative rule of probability

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

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) \times P(A_2) \times \ldots \times P(A_k)$$

Example K.5

Given the scenarios of previous example, 5 persons with diabetes are randomly selected.

(a) What is the probability that none of them receives any oral medication?

(b) What is the probability that all of them receives oral medication?

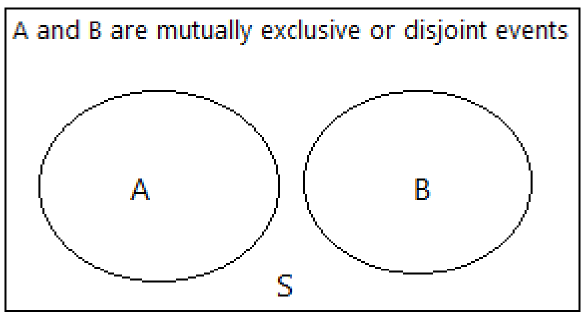
Solution

(a) $0.28 * 0.28 * 0.28 * 0.28 * 0.28 = 0.28^5 = 0.001721037$

(b) $0.72^5 = 0.1934918$

3.5 Addition Law of Probability

(a) **Disjoint events**: Let A and B be two disjoint or mutually exclusive events.



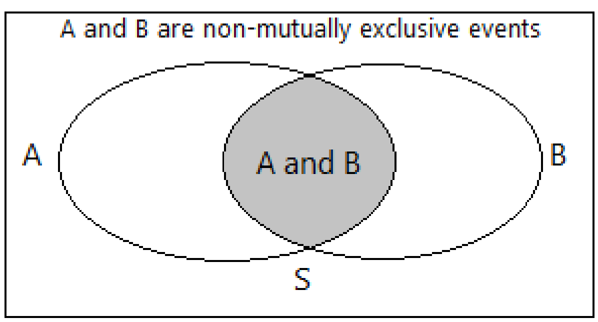
Then,

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

Or, expressed equivalently,

$$P(A \text{ or } B) = P(A) + P(B)$$

(b) **General or non-disjoint events**: Let A and B be two non-disjoint events.



Then,

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

Or, expressed equivalently,

$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$

(c) **Independent events:** Let A and B be two independent events. Then,

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

Or, $$P(A \cup B) = P(A) + P(B)[1 - P(A)]$$

Or, $$P(A \cup B) = P(B) + P(A)[1 - P(B)]$$

Example K.6

Table below provides the percentage distribution of blood type in the USA by different ethnic groups.

	Ethnic Group			
Blood Type	Caucasians	African-American	Hispanic	Asian
O	45%	51%	57%	40%
A	40%	26%	31%	28%
B	11%	19%	10%	25%
AB	4%	4%	2%	7%
Total	100%	100%	100%	100%

Source: 2018 American Red Cross, Blood Types and the Populations
http://www.redcrossblood.org/learn-about-blood/blood-types;

An African-American is selected at random.

(a) Find the probability that the randomly chosen African-American has either blood type O or B?

(b) A person with blood type A can receive blood transfusions from people with blood types O and A. What is the probability that a randomly chosen African-American can donate blood to persons with blood type A?

(c) A person with blood type AB can receive blood transfusions from people with blood types O, A, B or AB. What is the probability that a randomly chosen African-American can donate blood to persons with blood type AB?

Solution

(a) $P(O \cup B) = P(O) + P(B) = 0.51 + 0.19 = 0.70$

(b) $P(O\ or\ A) = 0.51 + 0.26 = 0.77$

(c) $P(O\ or\ A\ or\ B\ or\ AB) = 0.51 + 0.26 + 0.19 + 0.04 = 1$

Example K.7

Among those diagnosed with diabetes in the United States, 26% is treated by insulin, 72% is treated by oral medication and 14% is treated by both insulin and oral medication.

A person is randomly selected from those diagnosed with diabetes.

(a) Find the probability that the person is treated by insulin or oral medication.

(b) Find the probability that the person is treated by neither insulin nor oral medication.

(c) Person treated by only oral medication.

Solution

Let us define

I: the person is treated by insulin

O: the person is treated by oral medication

Then

(a) $P(I\ or\ O) = P(I) + P(O) - P(I\ and\ O) = 0.26 + 0.72 - 0.14 = 0.84$

(b) Treated by neither insulin nor oral medication is the event (not "$I\ or\ O$") and hence we have $P(not\ \text{"}I\ or\ O\text{"}) = 1 - 0.84 = 0.16$.

(c) $P(O\ only) = P(O) - P(O\ and\ I) = 0.72 - 0.14 = 0.58$

Example 3.16 TBP 49

Two doctors have the following results with the diagnosis of sexually transmitted disease:

$$P(A^+) = 0.1, P(B^+) = 0.17, P(A^+ \cap B^+) = 0.08$$

(a) Find the probability that either of the doctors has a positive diagnosis, i.e., compute the $P(A^+ \cup B^+) = 0.1 + 0.17 - 0.08 = ?$

(b) Is the event of diagnosis of two doctors independent?

Solution

(a) $P(A^+ \cup B^+) = P(A^+) + P(A^+) - P(A^+ \cap B^+) = 0.1 + 0.17 - 0.08 = 0.19$

(b) $P(A^+) \times P(B^+) = 0.1 * 0.17 = 0.017 \neq P(A^+ \cap B^+) = 0.08$

So, the diagnoses of two doctors are not independent.

Example K.

3.6 Conditional Probability

Let us define two events

A: a randomly selected person is a lung cancer patient.

B: a randomly selected person is a smoker.

Then, the event "$A|B$" stands for the event A given that B has occurred. Then, the probability of the event $A|B$, denoted by $P(A|B)$, refers to the probability of those who have lung cancer among smokers.

Definition: $P(A|B)$

The conditional probability of A given B is

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

In a similar manner, the conditional probability of B given A is given by

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

Remark: It follows that the conditional probability $P(A|B)$ can be computed using the formula $P(A|B) = \frac{n(A \cap B)}{n(B)}$ directly using the reduced sample space B.

Example 3.9

Suppose two doctors, A and B, test all patients coming to a clinic for syphilis. Define the following events:

$$A^{+} = \{doctor\ A\ makes\ a\ positive\ diagnosis\}$$

$$B^{+} = \{doctor\ B\ makes\ a\ positive\ diagnosis\}$$

Given $P(A^{+}) = 0.10$, $P(B^{+}) = 0.17$ and $P(A^{+} \cap B^{+}) = 0.08$.

(a) What is the probability that doctor B makes a positive diagnosis given that A makes a positive diagnosis? In other words, compute $P(B^{+}|A^{+})$.

(b) Compute $P(A^{+}|B^{+})$.

Solution

$P(B^{+}|A^{+}) = \frac{P(A^{+} \cap B^{+})}{P(A^{+})} = \frac{0.08}{0.1} = 0.8.$

$P(A^{+}|B^{+}) = \frac{P(A^{+} \cap B^{+})}{P(B^{+})} = \frac{0.08}{0.17} = 0.47$

Predictive value positive (PV+)

The predictive value positive (PV+) of a screening test is the probability that a person truly has a disease given that the test is positive. That is,

$$PV+= Pr(D^{+}|T^{+})$$

What does the PV+ imply?

How worried a subject with the positive test be?

Predictive value negative (PV-)

The predictive value negative (PV-) of a screening test is the probability that a person truly does not have a disease given that the test is negative. That is,

$$PV-= Pr(D^{-}|T^{-})$$

What does the PV- imply?

How reassured a subject with the negative test be?

Sensitivity

The sensitivity of a test is the probability that the test is positive given disease. That is

$$Sensitivity = Pr(T^{+}|D^{+})$$

Specificity

The specificity of a test is the probability that the test is negative given no disease. That is

$$Specificity = Pr(T^{-}|D^{-})$$