GREATER JAKARTA • BANDUNG • MALANG

STAT6171001
Basic Statistics

Probability
Session 4

Maria Seraphina Astriani
seraphina@binus.ac.id

People
Innovation
Excellence

Session Learning Outcomes

Upon completion of this session, students are expected to be able to

- LO 2. Analyze a problem by using the basic concept of descriptive and inferential statistics
- LO 3. Design a descriptive and inferential statistics solution to meet a given set of computing requirements in the context of computer science
- LO4. Produce descriptive and inferential statistics solutions

Topic

- Probability

Probability

Introduction

- In everyday life, we regularly generalize from limited sets of observations.
- One sip indicates that the batch of soup is too salty; dipping a toe in the swimming pool reassures us before taking the first plunge; etc.
- Conclusions that we'll encounter in inferential statistics, such as "95 percent confident" or "significant at the .05 level," are statements based on probabilities.

Concept of “Probability”

- Probability is an important and complex field of study
- Inferential statistics is built on the foundation of probability theory and has been remarkably successful in guiding opinion about the conclusions to be drawn from data.

Concept of "Probability"

- One conception of probability is drawn from the idea of **symmetrical outcomes**.

Example:

- The two possible outcomes of tossing a fair coin seem not to be distinguishable in any way that affects which side will land up or down.
- Therefore, the probability of heads is taken to be 1/2, as is the probability of tails.
- In general, if there are N symmetrical outcomes, the probability of any given one of them occurring is taken to be 1/N.
- Thus, if a six-sided die is rolled, the probability of any one of the six sides coming up is 1/6.

Concept of “Probability”

- Probabilities can also be thought of in terms of **relative frequencies**.

Example:

- If we tossed a coin millions of times, we would expect the proportion of tosses that came up heads to be pretty close to 1/2.
- As the number of tosses increases, the proportion of heads approaches 1/2. Therefore, we can say that the probability of a head is 1/2.

Probability of a Single Event

- If you roll a six-sided die, there are six possible outcomes, and each of these outcomes is equally likely.
- What is the probability that either a one or a six will come up?
- The two outcomes about which we are concerned (a one or a six coming up) are called favorable outcomes.

$$\text{Probability} = \frac{\text{Possible outcomes}}{\text{Total outcomes}}$$

- In this case there are two favorable outcomes and six possible outcomes.
- So, the probability of throwing either a one or six is 1/3

Probability of a Single Event

- The previous formula applies to many games of chance.
- For example, what is the probability that a card drawn at random from a deck of playing cards will be an ace?
- Since the deck has four aces, there are four favorable outcomes; since the deck has 52 cards, there are 52 possible outcomes.
- The probability is therefore 4/52 = 1/13.

- Here is a more complex example. You throw 2 dice. What is the probability that the sum of the two dice will be 6?

Probability of a Single Event

- The probability is 5/36.
- If you know the probability of an event occurring, it is easy to compute the probability that the event does not occur.
- The probability that the total is not 6 is 1 - 5/36 = 31/36

Table 1. 36 possible outcomes.

Die 1	Die 2	Total	Die 1	Die 2	Total	Die 1	Die 2	Total
1	1	2	3	1	4	5	1	6
1	2	3	3	2	5	5	2	7
1	3	4	3	3	6	5	3	8
1	4	5	3	4	7	5	4	9
1	5	6	3	5	8	5	5	10
1	6	7	3	6	9	5	6	11
2	1	3	4	1	5	6	1	7
2	2	4	4	2	6	6	2	8
2	3	5	4	3	7	6	3	9
2	4	6	4	4	8	6	4	10
2	5	7	4	5	9	6	5	11
2	6	8	4	6	10	6	6	12

Probability of Two (or more) Independent Events

- Events A and B are independent events if the probability of Event B occurring is the same whether Event A occurs.

Example:

- A fair coin is tossed two times.
- The probability that a head comes up on the second toss is 1/2 regardless of whether a head came up on the first toss.
- The two events are (1) first toss is a head and (2) second toss is a head.
- So, these events are independent.

Probability of Two (or more) Independent Events

Example:

- Consider the two events (1) “It will rain tomorrow in Houston” and (2) “It will rain tomorrow in Galveston” (a city near Houston).
- These events are not independent because it is more likely that it will rain in Galveston on days it rains in Houston than on days it does not.

Probability of A and B

- If events A and B are independent, then the probability of both A and B occurring is:

  `P(A and B) = P(A) x P(B)`

- where `P(A and B)` is the probability of events A and B both occurring, `P(A)` is the probability of event A occurring, and `P(B)` is the probability of event B occurring.
- If you flip two coins, what is the probability that it will come up both heads?

  `1/2 x 1/2 = 1/4`

Probability of A and B

- Another example, if you flip a coin and roll a six-sided die, what is the probability that the coin comes up heads and the die comes up 1?
- Since the two events are independent, the probability is simply the probability of a head (which is 1/2) times the probability of the die coming up 1 (which is 1/6).
- Therefore, the probability of both events occurring is

  1/2 x 1/6 = 1/12

Probability of A or B

- If Events A and B are dependent, the probability that either Event A or Event B occurs is:

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

- In this discussion, when we say "A or B occurs" we include three possibilities:
  1. A occurs and B does not occur
  2. B occurs and A does not occur
  3. Both A and B occur

Probability of A or B

Examples:

- If you flip a coin two times, what is the probability that you will get a head on the first flip or a head on the second flip (or both)?
- Letting Event A be a head on the first flip and Event B be a head on the second flip, then

$$P(A) = 1/2, P(B) = 1/2, \text{ and } P(A \text{ and } B) = 1/4$$

Therefore,

$$P(A \text{ or } B) = 1/2 + 1/2 - 1/4 = 3/4$$

Conditional Probabilities

- Often it is required to compute the probability of an event given that another event has occurred.

Example:

- What is the probability that two cards drawn at random from a deck of playing cards will both be aces?
- It might seem that you could use the formula for the probability of two independent events and simply multiply 4/52 x 4/52 = 1/169.
- This would be **incorrect**, because the two events are not independent.
- If the first card drawn is an ace, then the probability that the second card is also an ace would be lower because there would only be three aces left in the deck

Conditional Probabilities

- Once the first card chosen is an ace, the probability that the second card chosen is also an ace is called the conditional probability of drawing an ace.
- In this case, the “condition” is that the first card is an ace.

```
P(ace on second draw | an ace on the first draw)
```

- “The probability that an ace is drawn on the second draw given that an ace was drawn on the first draw.”

Conditional Probabilities

- Since after an ace is drawn on the first draw, there are 3 aces out of 51 total cards left. This means that the probability that one of these aces will be drawn is 3/51 = 1/17.
- If Events A and B are not independent, then

  `P(A and B) = P(A) x P(B|A)`
- Applying this to the problem of two aces, the probability of drawing two aces from a deck is

  `4/52 x 3/51 = 1/221`

Birthday Problem

- If there are 25 people in a room, what is the probability that at least two of them share the same birthday.
- If your first thought is that it is 25/365 = 0.068, you will be surprised to learn it is much higher than that.
- This problem requires the application of the sections on P(A and B) and conditional probability.
- This problem is best approached by asking what is the probability that no two people have the same birthday
- If we choose two people at random, what is the probability that they do not share a birthday?

Birthday Problem

- Of the 365 days on which the second person could have a birthday, 364 of them are different from the first person's birthday. Therefore, the probability is 364/365 (P2).
- Now define P3 as the probability that the 3rd person drawn doesn't share a birthday with anyone drawn previously given that there are no previous birthday matches.
- P3 is therefore a conditional probability. Therefore P3 = 363/365.
- In like manner, P4 = 362/365, P5 = 361/365, and so on up to P25 = 341/365.
- Since P(A and B) = P(A)P(B), all we have to do is multiply P2, P3, P4 ...P25 together.
- The result is 0.431. Therefore, the probability of at least one match is 0.569.

Permutations and Combinations

- The topics covered are:
  1. counting the number of possible orders,
  2. counting using the multiplication rule,
  3. counting the number of permutations, and
  4. counting the number of combinations

Possible Orders

- Suppose you had a plate with three pieces of candy on it: one green, one yellow, and one red.
- Pick up these three pieces one at a time.
- In how many different orders can you pick up the pieces?

Table 1. Six Possible Orders.

Number	First	Second	Third
1	red	yellow	green
2	red	green	yellow
3	yellow	red	green
4	yellow	green	red
5	green	red	yellow
6	green	yellow	red

Number of orders = n!

Multiplication Rule

- A small restaurant whose menu has 3 soups, 6 entrées, and 4 desserts.
- How many possible meals are there?
- The answer is calculated by multiplying the numbers to get

$$3 \times 6 \times 4 = 72$$

Permutations

- Suppose that there were four pieces of candy (red, yellow, green, and brown) and you were only going to pick up exactly two pieces.
- How many ways are there of picking up two pieces?

Table 2. Twelve Possible Orders.

Number	First	Second
1	red	yellow
2	red	green
3	red	brown
4	yellow	red
5	yellow	green
6	yellow	brown
7	green	red
8	green	yellow
9	green	brown
10	brown	red
11	brown	yellow
12	brown	green

Permutations

- The general formula is:

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

- where ${}_{n}P_{r}$ is the number of permutations of n things taken r at a time

$$ {}_{4}P_{2} = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12 $$

Combinations

- In counting combinations, choosing red and then yellow is the same as choosing yellow and then red.
- Unlike permutations, order does not count.

$$_{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

For our example,

$$_{4}C_{2} = \frac{4!}{(4-2)!\,2!} = \frac{4 \text{ x } 3 \text{ x } 2 \text{ x } 1}{(2 \text{ x } 1)(2 \text{ x } 1)} = 6$$

Table 3. Six Combinations.

Number	First	Second
1	red	yellow
2	red	green
3	red	brown
x	yellow	red
4	yellow	green
5	yellow	brown
x	green	red
x	green	yellow
6	green	brown
x	brown	red
x	brown	yellow
x	brown	green

Binomial Distribution

- The binomial distribution model is an important probability model that is used when there are two possible outcomes (hence "binomial").

Table 1. Four Possible Outcomes.

Outcome	First Flip	Second Flip
1	Heads	Heads
2	Heads	Tails
3	Tails	Heads
4	Tails	Tails

Binomial Distribution

Example:

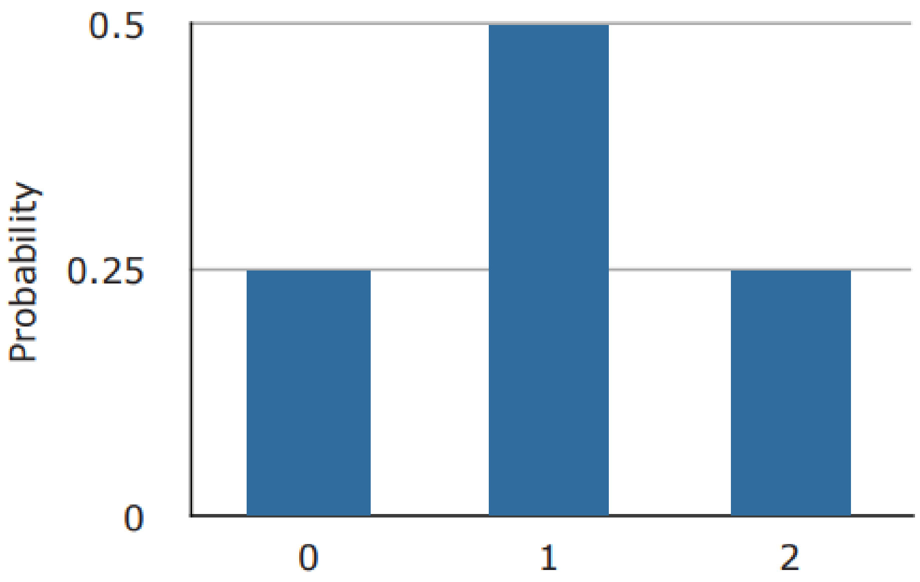
- The four possible outcomes that could occur if you flipped a coin twice are listed in Table 1.
- Note that the four outcomes are equally likely: each has probability 1/4.
- The probability of a head on Flip 1 and a head on Flip 2 is the product of P(H) and P(H), which is 1/2 x 1/2 = 1/4.
- The same calculation applies to the probability of a head on Flip 1 and a tail on Flip 2. Each is 1/2 x 1/2 = 1/4.

Binomial Distribution

- The 4 possible outcomes can be classified in terms of the number of heads that come up.
- The number could be two (Outcome 1), one (Outcomes 2 and 3) or 0 (Outcome 4).
- Since two of the outcomes represent the case in which just one head appears in the two tosses, the probability of this event is equal to 1/4 + 1/4 = 1/2.

Table 2. Probabilities of Getting 0, 1, or 2 Heads.

Number of Heads	Probability
0	1/4
1	1/2
2	1/4



The Formula for Binomial Probabilities

- The binomial distribution consists of the probabilities of each of the possible numbers of successes on N trials for independent events that each have a probability of π (the Greek letter pi) of occurring.

$$P(x) = \frac{N!}{x!\,(N-X)!}\pi^{x}(1-\pi)^{N-x}$$

- where P(x) is the probability of x successes out of N trials, N is the number of trials, and π is the probability of success on a given trial

The Formula for Binomial Probabilities

- Applying this to the coin flip example

$$P(0) = \frac{2!}{0!\,(2-0)!}(.5^0)(1-.5)^{2-0} = \frac{2}{2}(1)(.25) = 0.25$$

$$P(1) = \frac{2!}{1!\,(2-1)!}(.5^1)(1-.5)^{2-1} = \frac{2}{1}(.5)(.5) = 0.50$$

$$P(2) = \frac{2!}{2!\,(2-2)!}(.5^2)(1-.5)^{2-2} = \frac{2}{2}(.25)(1) = 0.25$$

Cumulative Probabilities

- Toss a coin 12 times. What is the probability that we get from 0 to 3 heads?
- The answer is found by computing the probability of exactly 0 heads, exactly 1 head, exactly 2 heads, and exactly 3 heads.
- The probability of getting from 0 to 3 heads is then the sum of these probabilities.
- The probabilities are: 0.0002, 0.0029, 0.0161, and 0.0537.
- The sum of the probabilities is 0.073

Mean and Standard Deviation of Binomial Distributions

- In general, the mean of a binomial distribution with parameters N (the number of trials) and π (the probability of success on each trial) is:

$$\mu = N\pi$$

where μ is the mean of the binomial distribution.

- The variance of the binomial distribution is:

$$\sigma^2 = N\pi(1-\pi)$$

where σ^2 is the variance of the binomial distribution

Mean and Standard Deviation of Binomial Distributions

- The coin was tossed 12 times, so N = 12.
- A coin has a probability of 0.5 of coming up heads.
- Therefore, $\pi = 0.5$.
- The mean and variance can therefore be computed as follows:

$$\mu = N\pi = (12)(0.5) = 6$$

$$\sigma 2 = N\pi(1-\pi) = (12)(0.5)(1.0 - 0.5) = 3.0$$

Poisson Distribution

- A Poisson distribution is a discrete probability distribution, meaning that it gives the probability of a discrete (i.e., countable) outcome.
- For Poisson distributions, the discrete outcome is the number of times an event occurs.
- You can use a Poisson distribution to predict or explain the number of events occurring within a given interval of time or space.
- “Events” could be anything from disease cases to customer purchases to meteor strikes.
- The interval can be any specific amount of time or space, such as 10 days or 5 square inches.

Poisson Distribution

- Can be used to calculate the probabilities of various numbers of “successes” based on the mean number of successes.
- The term “success” doesn’t really mean success in the traditional positive sense; it just means that the outcome in question occurs.

$$p = \frac{e^{-\mu}\mu^{x}}{x!}$$

e is the base of natural logarithms (2.7183)
μ is the mean number of “successes”
x is the number of “successes” in question

Poisson Distribution

Example:

- Suppose you knew that the mean number of calls to a fire station on a weekday is 8.
- What is the probability that on a given weekday there would be 11 calls?

$$p = \frac{e^{-8} 8^{11}}{11!} = 0.072$$

since the mean is 8 and the question pertains to 11 fires

Multinomial Distribution

- The multinomial distribution is the generalization of the binomial distribution to the case of n repeated trials where there are more than two possible outcomes to each.

Multinomial Distribution

Example:

- Suppose that two chess players had played numerous games, and it was determined that the **probability** that **Player A would win is 0.40**, the probability that **Player B would win is 0.35**, and the probability that the game would **end in a draw is 0.25**.
- The multinomial distribution can be used to answer questions such as: “If these two chess players played 12 games, what is the probability that Player A would win 7 games, Player B would win 2 games, and the remaining 3 games would be drawn?”

Multinomial Distribution

- The following formula gives the probability of obtaining a specific set of outcomes when there are k possible outcomes for each event:

$$p = \frac{n!}{(n_1!)(n_2!)\ldots(n_k!)} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

- where:
  - p is the probability
  - n is the total number of events
  - n_1 is the number of times Outcome 1 occurs,
  - n_2 is the number of times Outcome 2 occurs,
  - n_k is the number of times Outcome k occurs,
  - p_1 is the probability of Outcome 1
  - p_2 is the probability of Outcome 2, and
  - p_k is the probability of Outcome k

Multinomial Distribution

- For the chess example,
  - $n = 12$ (12 games are played),
  - $n_1 = 7$ (number won by Player A),
  - $n_2 = 2$ (number won by Player B),
  - $n_3 = 3$ (the number drawn),
  - $p_1 = 0.40$ (probability Player A wins)
  - $p_2 = 0.35$ (probability Player B wins)
  - $p_3 = 0.25$ (probability of a draw)

$$p = \frac{12!}{(7!)(2!)(3!)} .40^{7} .35^{2} .25^{3} = 0.0248$$

Hypergeometric Distribution

- The hypergeometric distribution is used to calculate probabilities when sampling without replacement.
- Example, suppose you first randomly sample 1 card from a deck of 52.
- Then, without putting the card back in the deck you sample a second and then a third.
- Given this sampling procedure, what is the probability that exactly two of the sampled cards will be aces (4 of the 52 cards in the deck are aces).

Hypergeometric Distribution

- Formula based on the hypergeometric distribution

$$p = \frac{{}_{k}C_{x} \; {}_{(N-k)}C_{(n-x)}}{{}_{N}C_{n}}$$

- Where:
  - k is the number of “successes” in the population
  - x is the number of “successes” in the sample
  - N is the size of the population
  - n is the number sampled
  - p is the probability of obtaining exactly x successes
  - ${}_{k}C_{x}$ is the number of combinations of k things taken x at a time

Hypergeometric Distribution

- In this example, k = 4 (four aces in the deck), x = 2 (probability of getting two aces), N = 52 (52 cards in a deck), and n = 3 because 3 cards were sampled.
- Therefore,

$$p = \frac{{}_4C_2 \; {}_{(52-4)}C_{(3-2)}}{{}_{52}C_3}$$

$$p = \frac{\frac{4!}{2!2!}\,\frac{48!}{47!1!}}{\frac{52!}{49!3!}} = 0.013$$

Hypergeometric Distribution

- The mean and standard deviation of the hypergeometric distribution are:

$$\text{mean} = \frac{(n)(k)}{N}$$

$$\text{sd} = \sqrt{\frac{(n)(k)(N-k)(N-n)}{N^2(N-1)}}$$

Standard Deviation

- A standard deviation is a measure of how dispersed the data is in relation to the mean.
- Low, or small, standard deviation indicates data are clustered tightly around the mean, and high, or large, standard deviation indicates data are more spread out.

Base Rates

- Suppose that at your regular physical exam you test positive for Disease X.
- Although Disease X has only mild symptoms, you are concerned and ask your doctor about the accuracy of the test.
- It turns out that the test is 95% accurate.
- The probability that you have Disease X is therefore 0.95.
- However, the situation is not that simple

Base Rates

- For one thing, more information about the accuracy of the test is needed because there are two kinds of errors the test can make: **misses** and **false positives**.
- If you actually have Disease X and the test failed to detect it, that would be a miss.
- If you did not have Disease X and the test indicated you did, that would be a false positive.
- The miss and false positive rates are not necessarily the same.

Base Rates

- For example, suppose that the test accurately indicates the disease in 99% of the people who have it and accurately indicates no disease in 91% of the people who do not have it.
- So, the test has a miss rate of 0.01 (1%) and a false positive rate of 0.09 (9%).
- This might lead you to revise your judgment and conclude that your chance of having the disease is 0.91.
- This would not be correct since the probability depends on the proportion of people having the disease.
- This proportion is called the base rate.

Base Rates

- Assume that Disease X is a rare disease, and only 2% of people in your situation have it. What would happen if 1 million people were tested.
- Out of these 1 million people, 2% or 20,000 people would have the disease.
- Of these 20,000 with the disease, the test would accurately detect it in 99% of them → 19,800 cases would be accurately identified.
- Let's consider the 98% of the 1 million people (980,000) who do not have the disease.
- Since the false positive rate is 0.09, 9% of these 980,000 people will test positive for the disease. This is a total of 88,200 people incorrectly diagnosed.

Base Rates

- To sum up, 19,800 people who tested positive would actually have the disease and 88,200 people who tested positive would not have the disease.
- This means that of all those who tested positive, only

  19,800/(19,800 + 88,200) = 0.1833

  of them would have the disease.
- So, the probability that you have the disease is not 0.95, or 0.91, but only 0.1833.

Base Rates

- These results are summarized in Table 1.

Table 1. Diagnosing Disease X.

True Condition			
No Disease 980,000		Disease 20,000	
Test Result		Test Result	
Positive 88,200	Negative 891,800	Positive 19,800	Negative 200

Bayes' Theorem

- This same result can be obtained using Bayes' theorem.
- Bayes' theorem considers both the prior probability of an event and the diagnostic value of a test to determine the posterior probability of the event.

For the current example

- The event is that you have Disease X. Let's call this Event D.
- Since only 2% of people in your situation have Disease X, the prior probability of Event D is 0.02 or, more formally, $P(D) = 0.02$.
- If D' represents the probability that Event D is false, then

$$P(D') = 1 - P(D) = 0.98$$

Bayes' Theorem

- To define the diagnostic value of the test, we need to define another event: that you test positive for Disease X → Event T.
- The diagnostic value of the test depends on the probability you will test positive given that you have the disease, written as P(T|D), and the probability you test positive given that you do not have the disease, written as P(T|D')

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D')P(D')}$$

```
P(T|D)  = 0.99
P(T|D') = 0.09
P(D)    = 0.02
P(D')   = 0.98
```

Bayes' Theorem

$$P(D|T) = \frac{(0.99)(0.02)}{(0.99)(0.02) + (0.09)(0.98)} = 0.1833$$

Populations

- Any complete set of observations (or potential observations) may be characterized as a population.
- Accurate descriptions of populations specify the nature of the observations to be taken.
- Example: a population might be described as “SAT critical reading scores of currently enrolled students at Rutgers University.”

Populations

- Real Populations
- Hypothetical Populations

Samples

- Any subset of observations from a population may be characterized as a sample.
- In typical applications of inferential statistics, the sample size is small relative to the population size.
- Example: less than 1 percent of all U.S. worksites are included in the Bureau of Labor Statistics' monthly survey to estimate the rate of unemployment.

References

- Witte, R.S.&Witte, J.S. (2017). Statistics (11th ed.). Wiley. ISBN: 978-1119386056.
- Lane, D.M., Scott, D., Hebl, M., Guerra, R., Osherson, D.& Zimmer, H. (2003). Introduction to Statistics. Online edition at https://open.umn.edu/opentextbooks/textbooks/459
- Levine, D.M., Stephan, D.F. & Szabat, K.A. (2017). Statistics for Managers Using Microsoft Excel (8th ed.). Pearson. ISBN: 978-0134566672

Thank you