Chapter 6. Probability

Likan Zhan
Beijing Language and Culture University

2021-10-25

https://likan.info
zhanlikan@blcu.edu.cn

Table of Contents

Table of Contents

Inferential Statistics and Probability

Inferential Statistics and Probability

Inferential Statistics and Probability

- Research studies begin with a general question about an entire population, but the actual research is conducted using a sample.

Inferential Statistics and Probability

- Research studies begin with a general question about an entire population, but the actual research is conducted using a sample.
- The role of inferential statistics is to use the sample data as the basis for answering questions about the population.

Inferential Statistics and Probability

- Research studies begin with a general question about an entire population, but the actual research is conducted using a sample.
- The role of inferential statistics is to use the sample data as the basis for answering questions about the population.
- To accomplish this goal, inferential procedures are typically built around the concept of probability.

Inferential Statistics and Probability

- Research studies begin with a general question about an entire population, but the actual research is conducted using a sample.
- The role of inferential statistics is to use the sample data as the basis for answering questions about the population.
- To accomplish this goal, inferential procedures are typically built around the concept of probability.
- Specifically, the relationships between samples and populations are usually defined in terms of *probability*.

Inferential Statistics and Probability

Inferential Statistics and Probability

- Suppose, for example, *you are selecting a single marble from a jar that contains 50 black and 50 white marbles*.

Inferential Statistics and Probability

- Suppose, for example, *you are selecting a single marble from a jar that contains 50 black and 50 white marbles*.
- In this example, the jar of marbles is the population and the single marble to be selected is the sample.

Inferential Statistics and Probability

- Suppose, for example, *you are selecting a single marble from a jar that contains 50 black and 50 white marbles*.
- In this example, the jar of marbles is the population and the single marble to be selected is the sample.
- Although you cannot guarantee the exact outcome of your sample, it is possible to talk about the potential outcomes in terms of probabilities.

Inferential Statistics and Probability

- Suppose, for example, *you are selecting a single marble from a jar that contains 50 black and 50 white marbles*.
- In this example, the jar of marbles is the population and the single marble to be selected is the sample.
- Although you cannot guarantee the exact outcome of your sample, it is possible to talk about the potential outcomes in terms of probabilities.
- In this case, *you have a 50—50 chance of getting either color*.

Inferential Statistics and Probability

Inferential Statistics and Probability

- Now consider *another jar (population) that has 90 black and only 10 white marbles*.

Inferential Statistics and Probability

- Now consider *another jar (population) that has 90 black and only 10 white marbles*.
- Again, you cannot specify the exact outcome of a sample, but now you know that *the sample probably will be a black marble*.

Inferential Statistics and Probability

Inferential Statistics and Probability

- By knowing the makeup of a population, we can determine the probability of obtaining specific samples.

Inferential Statistics and Probability

- By knowing the makeup of a population, we can determine the probability of obtaining specific samples.
- In this way, *probability* gives us a connection between populations and samples, and this connection is the foundation for the inferential statistics to be presented in the chapters that follow.

Inferential Statistics and Probability

- By knowing the makeup of a population, we can determine the probability of obtaining specific samples.
- In this way, *probability* gives us a connection between populations and samples, and this connection is the foundation for the inferential statistics to be presented in the chapters that follow.
- You may have noticed that the preceding examples begin with a population and then use probability to describe the samples that could be obtained.

Inferential Statistics and Probability

- By knowing the makeup of a population, we can determine the probability of obtaining specific samples.
- In this way, *probability* gives us a connection between populations and samples, and this connection is the foundation for the inferential statistics to be presented in the chapters that follow.
- You may have noticed that the preceding examples begin with a population and then use probability to describe the samples that could be obtained.
- This is exactly *backward* from what we want to do with inferential statistics.

Inferential Statistics and Probability

Inferential Statistics and Probability

- Remember that the goal of inferential statistics is to begin with a sample and then answer a general question about the population. We reach this goal in a two-stage process.

Inferential Statistics and Probability

- Remember that the goal of inferential statistics is to begin with a sample and then answer a general question about the population. We reach this goal in a two-stage process.
- In the first stage, we develop probability as a bridge from populations to samples. This stage involves identifying the types of samples that probably would be obtained from a specific population.

Inferential Statistics and Probability

- Remember that the goal of inferential statistics is to begin with a sample and then answer a general question about the population. We reach this goal in a two-stage process.
- In the first stage, we develop probability as a bridge from populations to samples. This stage involves identifying the types of samples that probably would be obtained from a specific population.
- Once this bridge is established, we simply reverse the probability rules to allow us to move from samples to populations.

Inferential Statistics and Probability

Inferential Statistics and Probability

- The process of reversing the probability relationship can be demonstrated by considering again the two jars of marbles we looked at earlier:

Inferential Statistics and Probability

- The process of reversing the probability relationship can be demonstrated by considering again the two jars of marbles we looked at earlier:
- *Jar 1 has 50 black and 50 white marbles*;

Inferential Statistics and Probability

- The process of reversing the probability relationship can be demonstrated by considering again the two jars of marbles we looked at earlier:
- *Jar 1 has 50 black and 50 white marbles*;
- *jar 2 has 90 black and only 10 white marbles*.

Inferential Statistics and Probability

- The process of reversing the probability relationship can be demonstrated by considering again the two jars of marbles we looked at earlier:
- *Jar 1 has 50 black and 50 white marbles*;
- *jar 2 has 90 black and only 10 white marbles*.
- This time, suppose you are blindfolded when the sample is selected, so you do not know which jar is being used.

Inferential Statistics and Probability

- The process of reversing the probability relationship can be demonstrated by considering again the two jars of marbles we looked at earlier:
- *Jar 1 has 50 black and 50 white marbles*;
- *jar 2 has 90 black and only 10 white marbles*.
- This time, suppose you are blindfolded when the sample is selected, so you do not know which jar is being used.
- Your task is to look at the sample that you obtain and then decide which jar is most likely.

Inferential Statistics and Probability

Inferential Statistics and Probability

- If you select a sample of $n = 4$ marbles and all are black, which jar would you choose?

Inferential Statistics and Probability

- If you select a sample of $n = 4$ marbles and all are black, which jar would you choose?
- It should be clear that it would be relatively unlikely (low probability) to obtain this sample from jar 1; in four draws, you almost certainly would get at least 1 white marble.

Inferential Statistics and Probability

- If you select a sample of $n = 4$ marbles and all are black, which jar would you choose?
- It should be clear that it would be relatively unlikely (low probability) to obtain this sample from jar 1; in four draws, you almost certainly would get at least 1 white marble.
- On the other hand, this sample would have a high probability of coming from jar 2, where nearly all the marbles are black.

Inferential Statistics and Probability

- If you select a sample of $n = 4$ marbles and all are black, which jar would you choose?
- It should be clear that it would be relatively unlikely (low probability) to obtain this sample from jar 1; in four draws, you almost certainly would get at least 1 white marble.
- On the other hand, this sample would have a high probability of coming from jar 2, where nearly all the marbles are black.
- Your decision therefore is that the sample probably came from jar 2.

Inferential Statistics and Probability

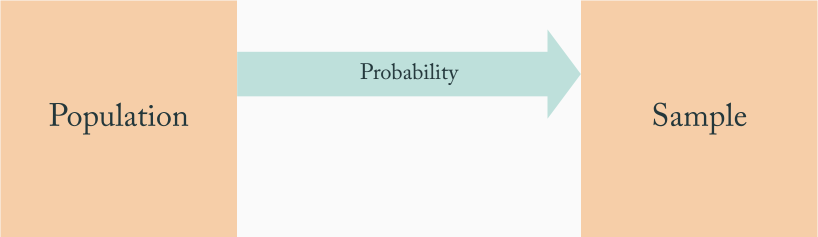
- If you select a sample of $n = 4$ marbles and all are black, which jar would you choose?
- It should be clear that it would be relatively unlikely (low probability) to obtain this sample from jar 1; in four draws, you almost certainly would get at least 1 white marble.
- On the other hand, this sample would have a high probability of coming from jar 2, where nearly all the marbles are black.
- Your decision therefore is that the sample probably came from jar 2.
- Note that you now are using the sample to make an inference about the population.

Inferential Statistics and Probability

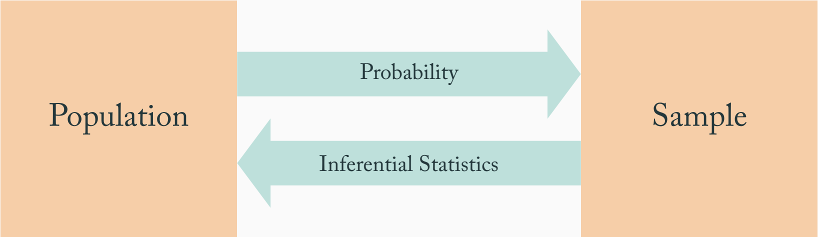
Population

Sample

Inferential Statistics and Probability



Inferential Statistics and Probability



Defining Probability

Defining Probability

Defining Probability

- Probability is a huge topic that extends far beyond the limits of introductory statistics, and we will not attempt to examine it all here.

Defining Probability

- Probability is a huge topic that extends far beyond the limits of introductory statistics, and we will not attempt to examine it all here.
- Instead, we concentrate on the few concepts and definitions that are needed for an introduction to inferential statistics.

Defining Probability

- Probability is a huge topic that extends far beyond the limits of introductory statistics, and we will not attempt to examine it all here.
- Instead, we concentrate on the few concepts and definitions that are needed for an introduction to inferential statistics.
- We begin with a relatively simple definition of probability.

Defining Probability

Defining Probability

- For a situation in which several different outcomes are possible, the **probability** (概率) for any specific outcome is defined as a fraction or a *proportion* (比率) of all the possible outcomes.

Defining Probability

- For a situation in which several different outcomes are possible, the **probability** (概率) for any specific outcome is defined as a fraction or a *proportion* (比率) of all the possible outcomes.
- If the possible outcomes are identified as A, B, C, D, and so on, then

$$\text{probability of A} = \frac{\text{number of outcomes classified as A}}{\text{total number of possible outcomes}}$$

Defining Probability

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

$$p(\text{the kings of hearts}) = \frac{1}{52}$$

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

$$p(\text{the kings of hearts}) = \frac{1}{52}$$

- The probability of selecting an ace is:

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

$$p(\text{the kings of hearts}) = \frac{1}{52}$$

- The probability of selecting an ace is:

$$p(\text{Ace}) = \frac{4}{52}$$

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

$$p(\text{the kings of hearts}) = \frac{1}{52}$$

- The probability of selecting an ace is:

$$p(\text{Ace}) = \frac{4}{52}$$

- The probability of selecting a spade card is:

Defining Probability

- For example, if you are selecting a card from a complete deck, there are 52 possible outcomes.
- The probability of selecting the king of hearts is:

$$p(\text{the kings of hearts}) = \frac{1}{52}$$

- The probability of selecting an ace is:

$$p(\text{Ace}) = \frac{4}{52}$$

- The probability of selecting a spade card is:

$$p(\text{spade}) = \frac{13}{52} = \frac{1}{4} = 0.25 = 25\%$$

Defining Probability

Defining Probability

- Or, if you are tossing a coin, the probability of getting heads is

Defining Probability

- Or, if you are tossing a coin, the probability of getting heads is

$$p(\text{heads}) = \frac{1}{2} = 0.5 = 50\%$$

Probability Values

Probability Values

Probability Values

- You also should note that all the possible probability values are contained in a limited range.

Probability Values

- You also should note that all the possible probability values are contained in a limited range.
- At one extreme, when an event never occurs, the probability is zero, or 0%.

Probability Values

- You also should note that all the possible probability values are contained in a limited range.
- At one extreme, when an event never occurs, the probability is zero, or 0%.
- At the other extreme, when an event always occurs, the probability is 1, or 100%.

Probability Values

- You also should note that all the possible probability values are contained in a limited range.
- At one extreme, when an event never occurs, the probability is zero, or 0%.
- At the other extreme, when an event always occurs, the probability is 1, or 100%.
- Thus, all probability values are contained in a range from 0-1.

Probability Values

Probability Values

- For example, suppose that you have a jar containing 10 white marbles.

Probability Values

- For example, suppose that you have a jar containing 10 white marbles.
- The probability of randomly selecting a black marble is

$$p(\text{black}) = \frac{0}{10} = 0$$

Probability Values

- For example, suppose that you have a jar containing 10 white marbles.
- The probability of randomly selecting a black marble is

$$p(\text{black}) = \frac{0}{10} = 0$$

- The probability of selecting a white marble is

$$p(\text{white}) = \frac{10}{10} = 1\%$$

Random Sampling

Random Sampling

Random Sampling

- For the preceding definition of probability to be accurate, it is necessary that the outcomes be obtained by a process called random sampling.

Random Sampling

- For the preceding definition of probability to be accurate, it is necessary that the outcomes be obtained by a process called random sampling.
- A **random sample** (随机取样) requires that each individual in the population has an *equal chance* of being selected.

Random Sampling

- For the preceding definition of probability to be accurate, it is necessary that the outcomes be obtained by a process called random sampling.
- A **random sample** (随机取样) requires that each individual in the population has an *equal chance* of being selected.
- A sample obtained by this process is called a **simple random sample** (简单随机样本).

Random Sampling

Random Sampling

- A second requirement, necessary for many statistical formulas, states that if more than one individual is being selected, the probabilities must *stay constant* from one selection to the next.

Random Sampling

- A second requirement, necessary for many statistical formulas, states that if more than one individual is being selected, the probabilities must *stay constant* from one selection to the next.
- Adding this second requirement produces what is called *independent random sampling*. The term *independent* refers to the fact that the probability of selecting any particular individual is independent of the individuals already selected for the sample.

Random Sampling

- A second requirement, necessary for many statistical formulas, states that if more than one individual is being selected, the probabilities must *stay constant* from one selection to the next.
- Adding this second requirement produces what is called *independent random sampling*. The term *independent* refers to the fact that the probability of selecting any particular individual is independent of the individuals already selected for the sample.
- For example, the probability that you will be selected is constant and does not change even when other individuals are selected before you are.

Random Sampling

Random Sampling

- An **independent random sample requires** (独立随机取样) that each individual has an equal chance of being selected and that the probability of being selected stays constant from one selection to the next if more than one individual is selected.

Random Sampling

Random Sampling

- Because independent random sample is usually a required component for most statistical applications, we will always assume that this is the sampling method being used.

Random Sampling

- Because independent random sample is usually a required component for most statistical applications, we will always assume that this is the sampling method being used.
- To simplify discussion, we will typically omit the word "independent" and simply refer to this sampling technique as random sampling.

Random Sampling

- Because independent random sample is usually a required component for most statistical applications, we will always assume that this is the sampling method being used.
- To simplify discussion, we will typically omit the word "independent" and simply refer to this sampling technique as random sampling.
- However, you should always assume that both requirements (*equal chance* and *constant probability*) are part of the process.

Random Sampling: Equal Chance

Random Sampling: Equal Chance

- Each of the two requirements for random sampling has some interesting consequences.

Random Sampling: Equal Chance

- Each of the two requirements for random sampling has some interesting consequences.
- The first assures that there is no bias in the selection process.

Random Sampling: Equal Chance

- Each of the two requirements for random sampling has some interesting consequences.
- The first assures that there is no bias in the selection process.
- The first requirement of random sampling prohibits you from applying the definition of probability to situations in which the possible outcomes are not equally likely.

Random Sampling: Constant Probability

Random Sampling: Constant Probability

- Consider, for example, the selection of $n = 2$ cards from a complete deck.

Random Sampling: Constant Probability

- Consider, for example, the selection of $n = 2$ cards from a complete deck.
- For the first draw, the probability of obtaining the jack of diamonds is

$$p(\text{jack of diamonds}) = \frac{1}{52} = 2\%$$

Random Sampling: Constant Probability

- Consider, for example, the selection of $n = 2$ cards from a complete deck.
- For the first draw, the probability of obtaining the jack of diamonds is

$$p(\text{jack of diamonds}) = \frac{1}{52} = 2\%$$

- After selecting one card for the sample, you are ready to draw the second card.

Random Sampling: Constant Probability

- Consider, for example, the selection of $n = 2$ cards from a complete deck.
- For the first draw, the probability of obtaining the jack of diamonds is

$$p(\text{jack of diamonds}) = \frac{1}{52} = 2\%$$

- After selecting one card for the sample, you are ready to draw the second card.
- What is the probability of obtaining the jack of diamonds this time?

Random Sampling: Constant Probability

Random Sampling: Constant Probability

- Assuming that you still are holding the first card, there are two possibilities:

Random Sampling: Constant Probability

- Assuming that you still are holding the first card, there are two possibilities:
- if the first card was not the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{1}{51} = 2\%$$

Random Sampling: Constant Probability

- Assuming that you still are holding the first card, there are two possibilities:
- if the first card was not the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{1}{51} = 2\%$$

- if the first card was the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{0}{51} = 0\%$$

Random Sampling: Constant Probability

- Assuming that you still are holding the first card, there are two possibilities:
- if the first card was not the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{1}{51} = 2\%$$

- if the first card was the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{0}{51} = 0\%$$

- In either case, the probability is different from its value for the first draw.

Random Sampling: Constant Probability

- Assuming that you still are holding the first card, there are two possibilities:
- if the first card was not the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{1}{51} = 2\%$$

- if the first card was the jack of diamonds

$$p(\text{jack of diamonds}) = \frac{0}{51} = 0\%$$

- In either case, the probability is different from its value for the first draw.
- This contradicts the requirement for random sampling, which says that the probability must stay constant.

Random Sampling: Constant Probability

Random Sampling: Constant Probability

- To keep the probabilities from changing from one selection to the next, it is necessary to return each individual to the population before you make the next selection. This process is called *sampling with replacement* (放回取样).

Random Sampling: Constant Probability

- To keep the probabilities from changing from one selection to the next, it is necessary to return each individual to the population before you make the next selection. This process is called *sampling with replacement* (放回取样).
- The second requirement for random samples (constant probability) demands that you sample with replacement.

Probability and Frequency Distributions

Probability and Frequency Distributions

Probability and Frequency Distributions

- We will use a very simple population that contains only $N = 10$ scores with values 1, 1, 2, 3, 3, 4, 4, 4, 5, 6.

Probability and Frequency Distributions

- We will use a very simple population that contains only $N = 10$ scores with values 1, 1, 2, 3, 3, 4, 4, 4, 5, 6.
- If you are taking a random sample of $n = 1$ score from this population, what is the probability of obtaining an individual with a score greater than 4?

Probability and Frequency Distributions

- We will use a very simple population that contains only $N = 10$ scores with values 1, 1, 2, 3, 3, 4, 4, 4, 5, 6.
- If you are taking a random sample of $n = 1$ score from this population, what is the probability of obtaining an individual with a score greater than 4?
- In probability notation,

$$p(X > 4) = ?$$

Probability and Frequency Distributions

Probability and Frequency Distributions

- The situations in which we are concerned with probability usually involve a population of scores that can be displayed in a frequency distribution graph.

Probability and Frequency Distributions

- The situations in which we are concerned with probability usually involve a population of scores that can be displayed in a frequency distribution graph.
- If you think of the graph as representing the entire population, then different portions of the graph represent different portions of the population.

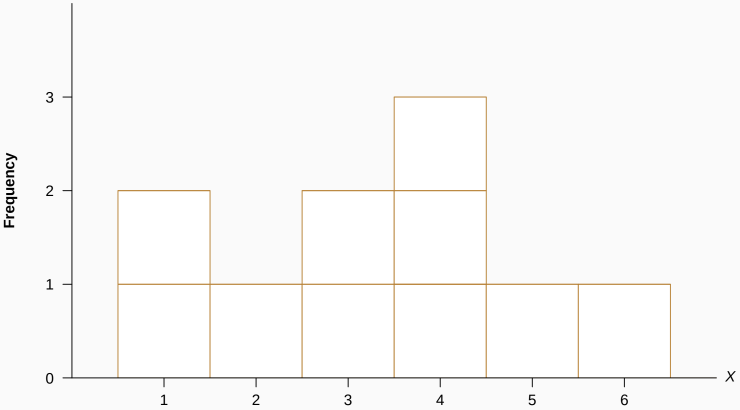
Probability and Frequency Distributions

- The situations in which we are concerned with probability usually involve a population of scores that can be displayed in a frequency distribution graph.
- If you think of the graph as representing the entire population, then different portions of the graph represent different portions of the population.
- Because probabilities and proportions are equivalent, a particular portion of the graph corresponds to a particular probability in the population.

Probability and Frequency Distributions

- The situations in which we are concerned with probability usually involve a population of scores that can be displayed in a frequency distribution graph.
- If you think of the graph as representing the entire population, then different portions of the graph represent different portions of the population.
- Because probabilities and proportions are equivalent, a particular portion of the graph corresponds to a particular probability in the population.
- Thus, whenever a population is presented in a frequency distribution graph, it will be possible to represent probabilities as proportions of the graph.

Probability and Frequency Distributions



Probability and Frequency Distributions

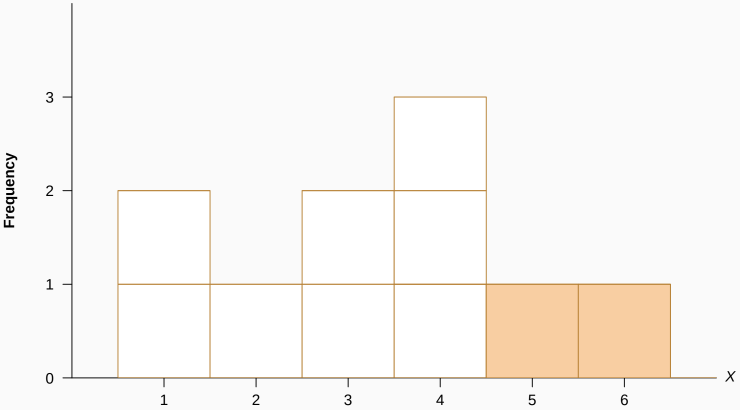


Table of Contents

The Normal Distribution

The normal distribution

The normal distribution

- The *normal distribution* (正态分布) is an example of a commonly occurring shape for population distributions.

The normal distribution

- The *normal distribution* (正态分布) is an example of a commonly occurring shape for population distributions.
- The exact shape of the normal distribution is specified by an equation relating each X value (score) with each Y value (frequency). The equation is

$$Y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

(π and e are mathematical constants)

The normal distribution

- The *normal distribution* (正态分布) is an example of a commonly occurring shape for population distributions.
- The exact shape of the normal distribution is specified by an equation relating each X value (score) with each Y value (frequency). The equation is

$$Y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

  (π and e are mathematical constants)
- In simpler terms, the normal distribution is symmetrical with a single mode in the middle.The frequency tapers off as you move farther from the middle in either direction.

The normal distribution

The normal distribution

- The normal shape can also be described by the proportions of area contained in each section of the distribution.

The normal distribution

- The normal shape can also be described by the proportions of area contained in each section of the distribution.
- A distribution is normal if and only if it has all the right proportions.

The normal distribution

The normal distribution

- There are two additional points to be made about the distribution.

The normal distribution

- There are two additional points to be made about the distribution.
- First, you should realize that the sections on the left side of the distribution have exactly the same areas as the corresponding sections on the right side because the normal distribution is symmetrical.

The normal distribution

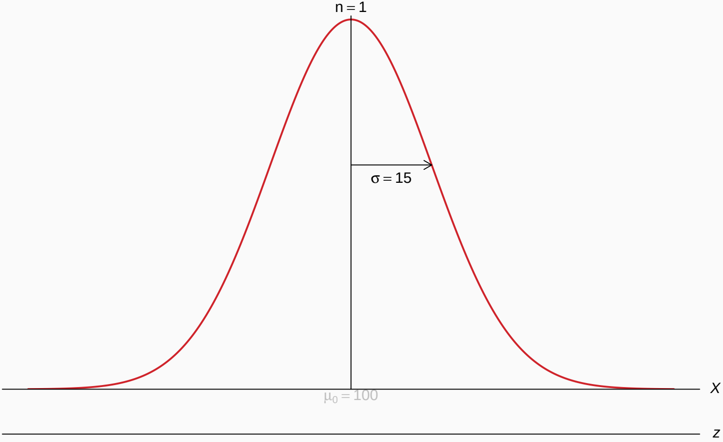
The normal distribution

- Second, because the locations in the distribution are identified by z-scores, the percentages shown in the figure apply to any normal distribution regardless of the values for the mean and the standard deviation.

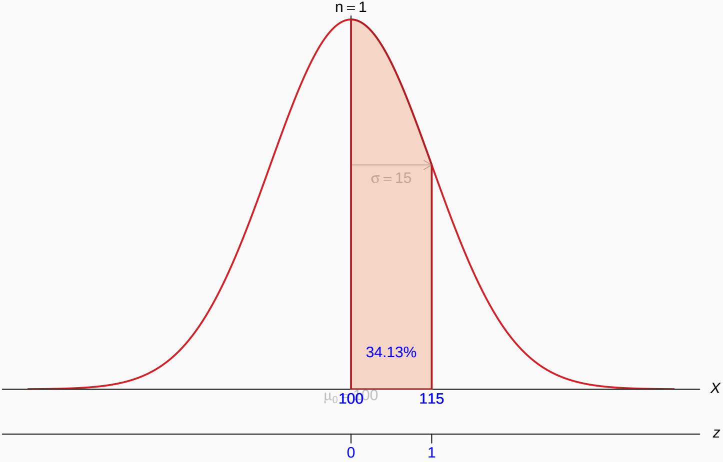
The normal distribution

- Second, because the locations in the distribution are identified by z-scores, the percentages shown in the figure apply to any normal distribution regardless of the values for the mean and the standard deviation.
- Because the normal distribution is a good model for many naturally occurring distributions and because this shape is guaranteed in some circumstances, we devote considerable attention to this particular distribution.

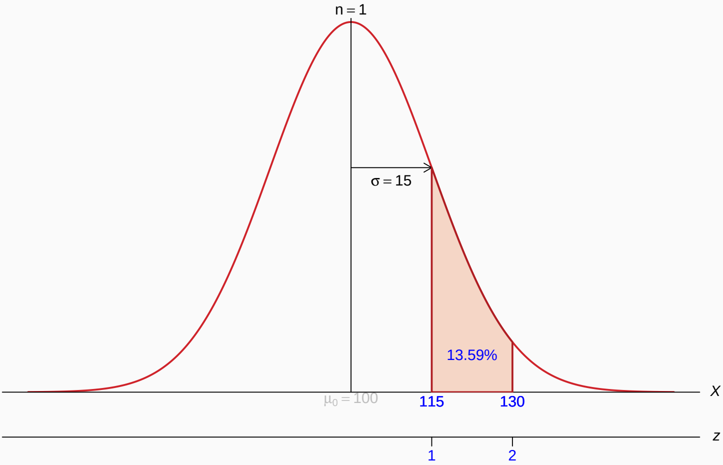
The normal distribution



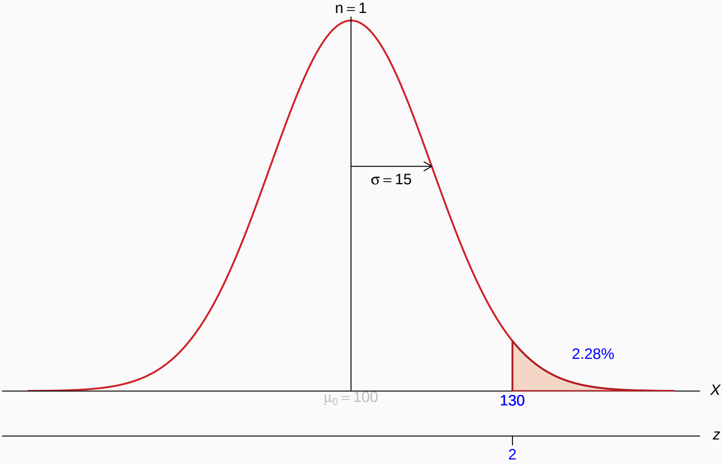
The normal distribution



The normal distribution after z-transformation



The normal distribution after z-transformation



The population distribution of SAT scores

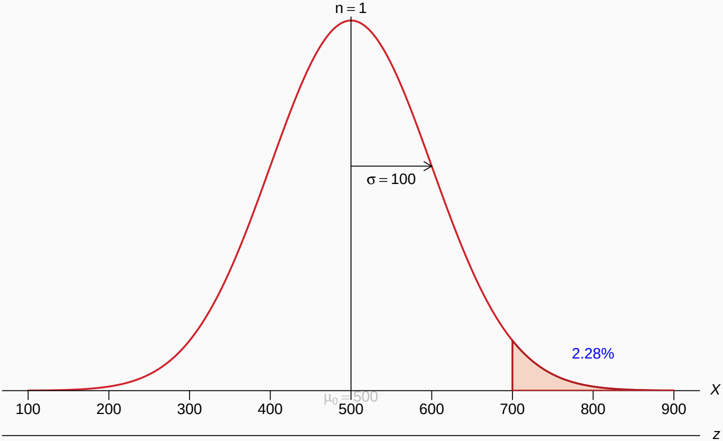
The population distribution of SAT scores

- The population distribution of SAT scores is normal with a mean of $\mu = 500$, a standard deviation of $\sigma = 100$.

The population distribution of SAT scores

- The population distribution of SAT scores is normal with a mean of $\mu = 500$, a standard deviation of $\sigma = 100$.
- What is the probability of randomly selecting an individual from this population who has an SAT score greater than 700?
  $p(X > 700) = ?$

The population distribution of SAT scores

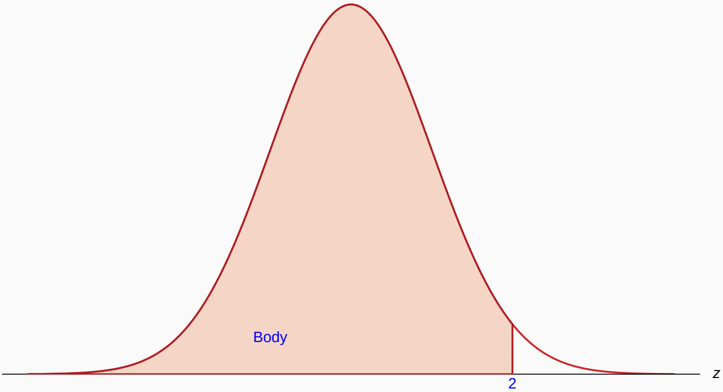


The Unit Normal Distribution

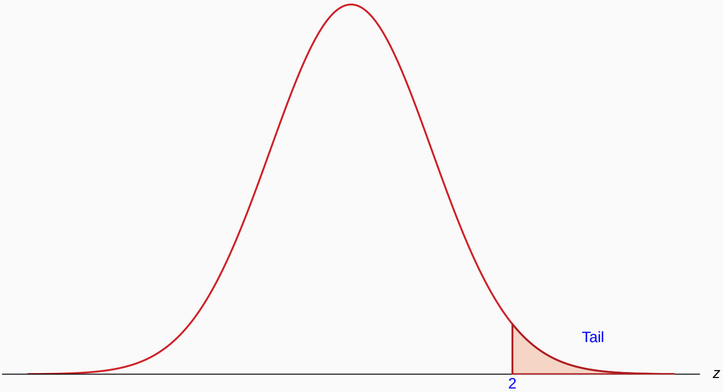
The Unit Normal Table

(A) z	(B) Proportion in body	(C) Proportion in tail	(D) Proportion between mean and z
0.00	.5000	.5000	.0000
0.01	.5040	.4960	.0040
0.02	.5080	.4920	.0080
0.03	.5120	.4880	.0120
0.21	.5832	.4168	.0832
0.22	.5871	.4129	.0871
0.23	.5910	.4090	.0910
0.24	.5948	.4052	.0948
0.25	.5987	.4013	.0987
0.26	.6026	.3974	.1026
0.27	.6064	.3936	.1064
0.28	.6103	.3897	.1103
0.29	.6141	.3859	.1141
0.30	.6179	.3821	.1179
0.31	.6217	.3783	.1217
0.32	.6255	.3745	.1255
0.33	.6293	.3707	.1293
0.34	.6331	.3669	.1331

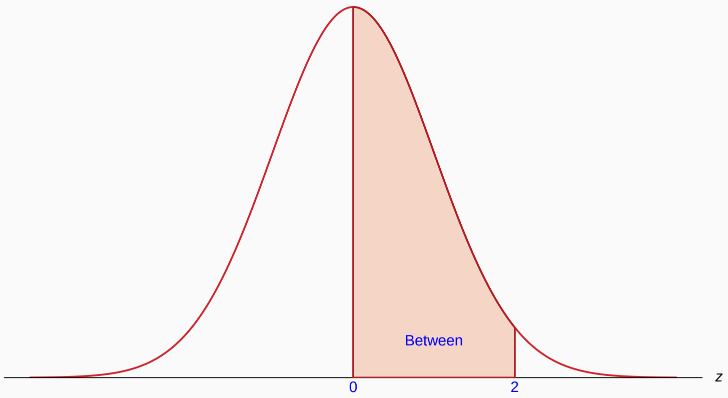
The Unit Normal Table: Body



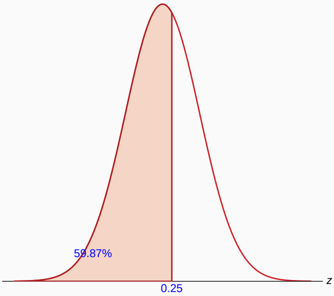
The Unit Normal Table: Tail

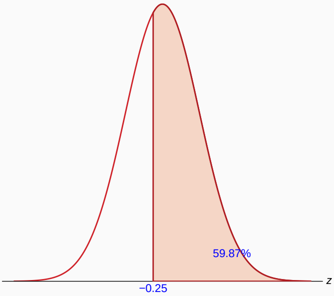


The Unit Normal Table: Between



The normal distribution is symmetrical





The Unit Normal Table

The Unit Normal Table

- The body always corresponds to the larger part of the distribution whether it is on the right-hand side or the left-hand side. Similarly, the tail is always the smaller section whether it is on the right or the left.

The Unit Normal Table

- The body always corresponds to the larger part of the distribution whether it is on the right-hand side or the left-hand side. Similarly, the tail is always the smaller section whether it is on the right or the left.
- Because the normal distribution is symmetrical, the proportions on the right-hand side are exactly the same as the corresponding proportions on the left-hand side.

The Unit Normal Table

- The body always corresponds to the larger part of the distribution whether it is on the right-hand side or the left-hand side. Similarly, the tail is always the smaller section whether it is on the right or the left.
- Because the normal distribution is symmetrical, the proportions on the right-hand side are exactly the same as the corresponding proportions on the left-hand side.
- Although the z-score values change signs (+ and -) from one side to the other, the proportions are always positive.

Probabilities, Proportions, and z-Scores

Probabilities, Proportions, and z-Scores

Probabilities, Proportions, and z-Scores

- The unit normal table lists relationships between z-score locations and proportions in a normal distribution.

Probabilities, Proportions, and z-Scores

- The unit normal table lists relationships between z-score locations and proportions in a normal distribution.
- For any z-score location, you can use the table to look up the corresponding proportions.

Probabilities, Proportions, and z-Scores

- The unit normal table lists relationships between z-score locations and proportions in a normal distribution.
- For any z-score location, you can use the table to look up the corresponding proportions.
- Similarly, if you know the proportions, you can use the table to find the specific z-score location.

Probabilities, Proportions, and z-Scores

- The unit normal table lists relationships between z-score locations and proportions in a normal distribution.
- For any z-score location, you can use the table to look up the corresponding proportions.
- Similarly, if you know the proportions, you can use the table to find the specific z-score location.
- Because we have defined probability as equivalent to proportion, you can also use the unit normal table to look up probabilities for normal distributions.

Probabilities, Proportions, and z-Scores

Function	Purpose	Example
rnorm (n, mean, sd)	Generate random numbers from normal distribution	*rnorm (100, 3, 1)* generates 100 numbers from a normal with mean=3 and sd=1
dnorm (x, mean, sd)	Probability Density Function (PDF)	*dnorm (0, 3, 1)* gives the density (height of the PDF) of the normal 0 with mean=3 and sd=1
pnorm (q, mean, sd)	Cumulative Distribution Function (CDF)	*pnorm (1.96, 0, 1)* gives the area under the standard normal curve to the left of 1.96, i.e. 0.975
qnorm (p, mean, sd)	Quantile Function - inverse of pnorm	*qnorm (0.975, 0, 1)* gives the value at which the CDF of the standard normal is .975, i.e. 1.96

Finding Proportions for z-Score Values

```
## pnorm(q, mean = 0, sd = 1,
##     lower.tail = TRUE, log.p = FALSE)
```

Finding Proportions for z-Score Values

Finding Proportions for z-Score Values

- What proportion of the normal distribution corresponds to z-score values greater than $z = +1.00$?

Finding Proportions for z-Score Values

- What proportion of the normal distribution corresponds to z-score values greater than $z = +1.00$?
-
```r
pnorm(1.00, lower.tail = FALSE)
## [1] 0.1586553
```

Finding Proportions for z-Score Values

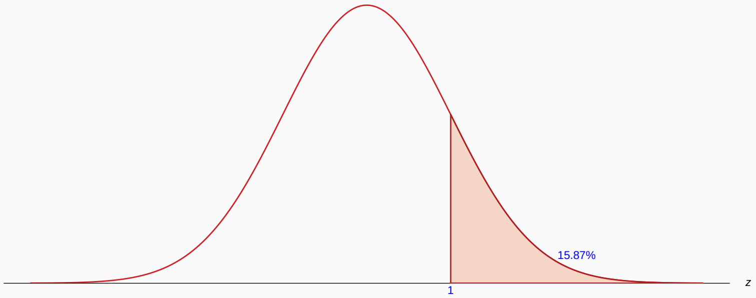
- What proportion of the normal distribution corresponds to z-score values greater than $z = +1.00$?
- ```
  pnorm(1.00, lower.tail = FALSE)
  ## [1] 0.1586553
  ```
- ```
  1 - pnorm(1.00, lower.tail = TRUE)
  ## [1] 0.1586553
  ```

Finding Proportions for z-Score Values

- What proportion of the normal distribution corresponds to z-score values greater than $z = +1.00$?

- ```
  pnorm(1.00, lower.tail = FALSE)

  ## [1] 0.1586553
  ```

- ```
  1 - pnorm(1.00, lower.tail = TRUE)

  ## [1] 0.1586553
  ```

- ```
  pnorm(-1.00, lower.tail = TRUE)

  ## [1] 0.1586553
  ```
  ```

Finding Proportions for z-Score Values

- What proportion of the normal distribution corresponds to z-score values greater than $z = +1.00$?

- ```
  pnorm(1.00, lower.tail = FALSE)
  ## [1] 0.1586553
  ```

- ```
  1 - pnorm(1.00, lower.tail = TRUE)
  ## [1] 0.1586553
  ```

- ```
  pnorm(-1.00, lower.tail = TRUE)
  ## [1] 0.1586553
  ```

- ```
  1 - pnorm(-1.00, lower.tail = FALSE)
  ## [1] 0.1586553
  ```

Finding Proportions for z-Score Values



Finding Proportions for z-Score Values

Finding Proportions for z-Score Values

- For a normal distribution, what is the probability of selecting a z-score less than $z = 1.50$? In symbols, $p(z < 1.50) = ?$

Finding Proportions for z-Score Values

- For a normal distribution, what is the probability of selecting a z-score less than $z = 1.50$? In symbols, $p(z < 1.50) = ?$

-
```
pnorm(1.50, lower.tail = TRUE)

## [1] 0.9331928
```

Finding Proportions for z-Score Values

- For a normal distribution, what is the probability of selecting a z-score less than $z = 1.50$? In symbols, $p(z < 1.50) =?$

- ```
  pnorm(1.50, lower.tail = TRUE)
  ```
  ```
  ## [1] 0.9331928
  ```

- what proportion of the normal distribution is contained in the tail beyond $z = -0.50$? That is, $p(z < -0.50)$
  ```

Finding Proportions for z-Score Values

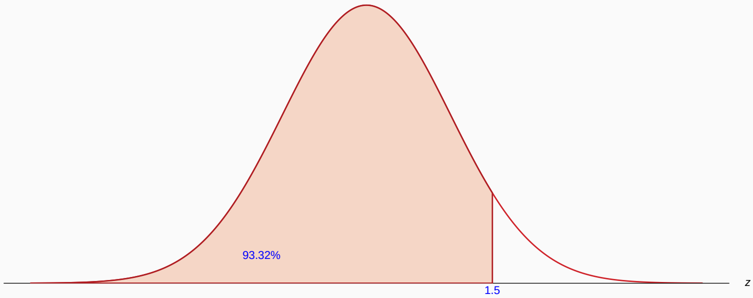
- For a normal distribution, what is the probability of selecting a z-score less than $z = 1.50$? In symbols, $p(z < 1.50) = ?$

- ```
  pnorm(1.50, lower.tail = TRUE)
  ## [1] 0.9331928
  ```

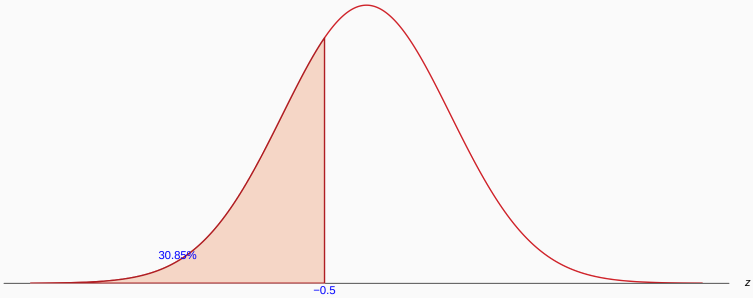
- what proportion of the normal distribution is contained in the tail beyond $z = -0.50$? That is, $p(z < -0.50)$

- ```
  pnorm(-0.50, lower.tail = TRUE)
  ## [1] 0.3085375
  ```

Finding Proportions for z-Score Values



Finding Proportions for z-Score Values



Finding Proportions for Specific z-Score Values

Finding Proportions for Specific z-Score Values

- $z > 0.80$,

Finding Proportions for Specific z-Score Values

- $z > 0.80, p = 0.2118554$

Finding Proportions for Specific z-Score Values

- $z > 0.80, p = 0.2118554$
- $z > -0.75,$

Finding Proportions for Specific z-Score Values

- $z > 0.80, p = 0.2118554$
- $z > -0.75, p = 0.7733726$

Finding the z-Score to Specific Proportions

```
## qnorm(p, mean = 0, sd = 1,
##     lower.tail = TRUE, log.p = FALSE)
```

Finding the z-Score to Specific Proportions

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score separates the top 10% from the remainder of the distribution?

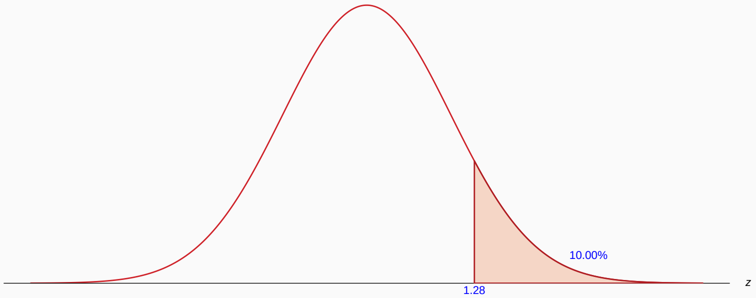
Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score separates the top 10% from the remainder of the distribution?
- ```
  qnorm(0.1, lower.tail = FALSE)
  ## [1] 1.281552
  ```
  ```

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score separates the top 10% from the remainder of the distribution?
- ```
  qnorm(0.1, lower.tail = FALSE)
  ## [1] 1.281552
  ```
- ```
  - qnorm(0.1, lower.tail = TRUE)
  ## [1] 1.281552
  ```

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score separates the top 10% from the remainder of the distribution?
- ```
  qnorm(0.1, lower.tail = FALSE)
  ## [1] 1.281552
  ```
- ```
  - qnorm(0.1, lower.tail = TRUE)
  ## [1] 1.281552
  ```
- ```
  qnorm(1 - 0.1, lower.tail = TRUE)
  ## [1] 1.281552
  ```
  ```

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score separates the top 10% from the remainder of the distribution?
- ```r
  qnorm(0.1, lower.tail = FALSE)
  ## [1] 1.281552
  ```
- ```r
  - qnorm(0.1, lower.tail = TRUE)
  ## [1] 1.281552
  ```
- ```r
  qnorm(1 - 0.1, lower.tail = TRUE)
  ## [1] 1.281552
  ```
- ```r
  - qnorm(1 - 0.1, lower.tail = FALSE)
  ## [1] 1.281552
  ```

Finding the z-Score to Specific Proportions



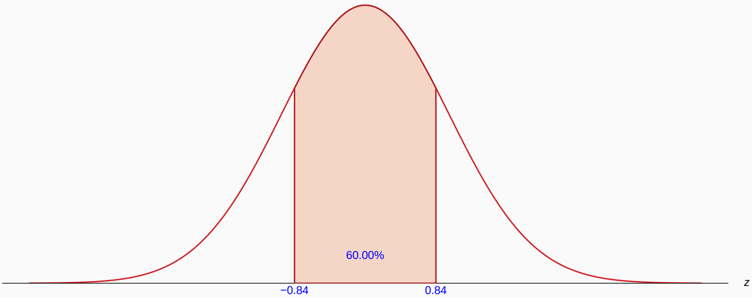
Finding the z-Score to Specific Proportions

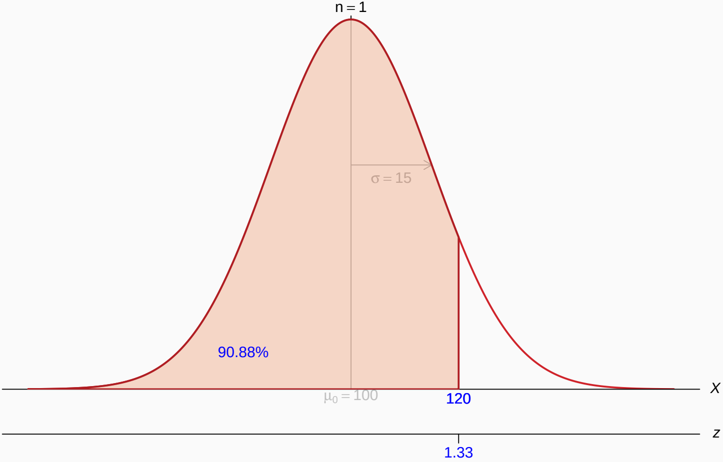
Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score values form the boundaries that separate the middle 60% of the distribution from the rest of the scores?

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score values form the boundaries that separate the middle 60% of the distribution from the rest of the scores?

- ```r
  qnorm(0.5 - 0.6/2, lower.tail = FALSE)
  ## [1] 0.8416212
  ```
  ```

Finding the z-Score to Specific Proportions

- For a normal distribution, what z-score values form the boundaries that separate the middle 60% of the distribution from the rest of the scores?

- ```
  qnorm(0.5 - 0.6/2, lower.tail = FALSE)
  ## [1] 0.8416212
  ```

- ```
  qnorm(0.5 + 0.6/2, lower.tail = FALSE)
  ## [1] -0.8416212
  ```

Finding the z-Score to Specific Proportions



Probabilities from a Normal Distribution

Probabilities from a Normal Distribution

Probabilities from a Normal Distribution

- In the preceding section, we used the unit normal table to find probabilities and proportions corresponding to specific z-score values.

Probabilities from a Normal Distribution

- In the preceding section, we used the unit normal table to find probabilities and proportions corresponding to specific z-score values.
- In most situations, however, it is necessary to find probabilities for specific X values.

Probabilities from a Normal Distribution

Probabilities from a Normal Distribution

- To answer probability questions about scores (X values) from a normal distribution, you must use the following two-step procedure:

Probabilities from a Normal Distribution

- To answer probability questions about scores (X values) from a normal distribution, you must use the following two-step procedure:
    1. Transform the X values into z-scores.

Probabilities from a Normal Distribution

- To answer probability questions about scores (X values) from a normal distribution, you must use the following two-step procedure:
  1. Transform the X values into z-scores.
  2. Use the unit normal table to look up the proportions corresponding to the z-score values.

Probabilities from a Normal Distribution

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$.

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$.
- Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$.
- Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?
- The first step is to change the X values into z-scores.

```
z <- (120 - 100) / 15
```

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$.
- Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?
- The first step is to change the X values into z-scores.

```
z <- (120 - 100) / 15
```

- Next, look up the z-score value in the unit normal table.

```
pnorm(z, lower.tail =  TRUE)

## [1] 0.9087888
```

Probabilities from a Normal Distribution



Probabilities from a Normal Distribution

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$. Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$. Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?

- ```
  ## pnorm(q, mean = 0, sd = 1,
  ##     lower.tail = TRUE, log.p = FALSE)
  ```
  ```

Probabilities from a Normal Distribution

- It is known that IQ scores form a normal distribution with $\mu = 100$, and $\sigma = 15$. Given this information, what is the probability of randomly selecting an individual with an IQ score less than 120?

- ```
  ## pnorm(q, mean = 0, sd = 1,
  ##      lower.tail = TRUE, log.p = FALSE)
  ```

- ```
  pnorm(120, mean = 100, sd = 15,
        lower.tail =  TRUE)
  ## [1] 0.9087888
  ```

Find Proportions Located between Two Scores

Find Proportions Located between Two Scores

- The highway department conducted a study measuring driving speeds on a local section of interstate highway.

Find Proportions Located between Two Scores

- The highway department conducted a study measuring driving speeds on a local section of interstate highway.
- They found an average speed of $\mu = 58$ miles per hour with a standard deviation of $\sigma = 10$. The distribution was approximately normal.

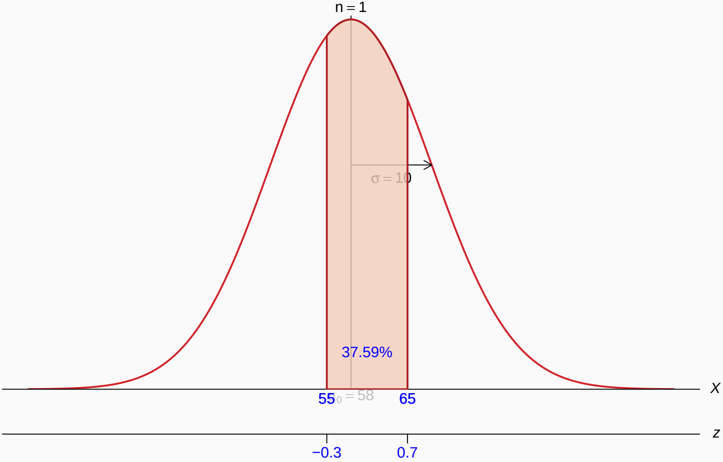
Find Proportions Located between Two Scores

- The highway department conducted a study measuring driving speeds on a local section of interstate highway.
- They found an average speed of $\mu = 58$ miles per hour with a standard deviation of $\sigma = 10$. The distribution was approximately normal.
- Given this information, what proportion of the cars are traveling between 55 and 65 miles per hour?

Find Proportions Located between Two Scores

- ```r
  (p1 <- pnorm(65, mean = 58, sd = 10))
  ## [1] 0.7580363
  ```
- ```r
  (p2 <- pnorm(55, mean = 58, sd = 10))
  ## [1] 0.3820886
  ```
- ```r
  p1 - p2
  ## [1] 0.3759478
  ```
  ```

Find Proportions Located between Two Scores



Find Proportions Located between Two Scores

Find Proportions Located between Two Scores

- Using the same distribution of driving speeds from the previous example, what proportion of cars is traveling between 65 and 75 miles per hour?

Find Proportions Located between Two Scores

- Using the same distribution of driving speeds from the previous example, what proportion of cars is traveling between 65 and 75 miles per hour?

-
```r
(p1 <- pnorm(75, mean = 58, sd = 10))
```
```
## [1] 0.9554345
```

Find Proportions Located between Two Scores

- Using the same distribution of driving speeds from the previous example, what proportion of cars is traveling between 65 and 75 miles per hour?

- ```
  (p1 <- pnorm(75, mean = 58, sd = 10))
  ## [1] 0.9554345
  ```

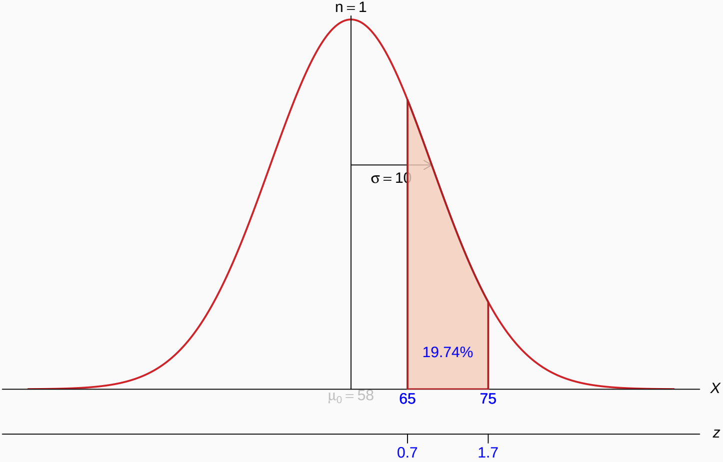
- ```
  (p2 <- pnorm(65, mean = 58, sd = 10))
  ## [1] 0.7580363
  ```

Find Proportions Located between Two Scores

- Using the same distribution of driving speeds from the previous example, what proportion of cars is traveling between 65 and 75 miles per hour?

- ```
  (p1 <- pnorm(75, mean = 58, sd = 10))
  ## [1] 0.9554345
  ```

- ```
  (p2 <- pnorm(65, mean = 58, sd = 10))
  ## [1] 0.7580363
  ```

- ```
  p1 - p2
  ## [1] 0.1973982
  ```
  ```

Find Proportions Located between Two Scores



Find Proportions Located between Two Scores

Find Proportions Located between Two Scores

- For a normal distribution with $\mu = 60$ and a standard deviation of $\sigma = 12$, find each probability requested.

Find Proportions Located between Two Scores

- For a normal distribution with $\mu = 60$ and a standard deviation of $\sigma = 12$, find each probability requested.
- $p(X > 66)$

```
pnorm(66, mean = 60, sd = 12, lower.tail = FALSE)
## [1] 0.3085375
```

Find Proportions Located between Two Scores

- For a normal distribution with $\mu = 60$ and a standard deviation of $\sigma = 12$, find each probability requested.
- $p(X > 66)$

```
pnorm(66, mean = 60, sd = 12, lower.tail = FALSE)
## [1] 0.3085375
```

- $p(48 < X < 72)$

```
p1 <- pnorm(72, mean = 60, sd = 12)
p2 <- pnorm(48, mean = 60, sd = 12)
p1 - p2
## [1] 0.6826895
```

Find Scores Corresponding to Specific Proportions

Find Scores Corresponding to Specific Proportions

- The U.S. Census Bureau (2005) reports that Americans spend an average of $\mu = 24.3$ minutes commuting to work each day.

Find Scores Corresponding to Specific Proportions

- The U.S. Census Bureau (2005) reports that Americans spend an average of $\mu = 24.3$ minutes commuting to work each day.
- Assuming that the distribution of commuting times is normal with a standard deviation of $\sigma = 10$ minutes, how much time do you have to spend commuting each day to be in the highest 10% nationwide?

Find Scores Corresponding to Specific Proportions

- The U.S. Census Bureau (2005) reports that Americans spend an average of $\mu = 24.3$ minutes commuting to work each day.
- Assuming that the distribution of commuting times is normal with a standard deviation of $\sigma = 10$ minutes, how much time do you have to spend commuting each day to be in the highest 10% nationwide?
- ```
  ## qnorm(p, mean = 0, sd = 1,
  ##      lower.tail = TRUE, log.p = FALSE)
  ```
  ```

Find Scores Corresponding to Specific Proportions

Find Scores Corresponding to Specific Proportions

- ```r
  qnorm(0.1, 24.3, 10, lower.tail = FALSE)
  ```

  ```
  ## [1] 37.11552
  ```
  ```

Find Scores Corresponding to Specific Proportions

- ```r
  qnorm(0.1, 24.3, 10, lower.tail = FALSE)
  ## [1] 37.11552
  ```
- ```r
  qnorm(1 - 0.1, 24.3, 10, lower.tail = TRUE)
  ## [1] 37.11552
  ```

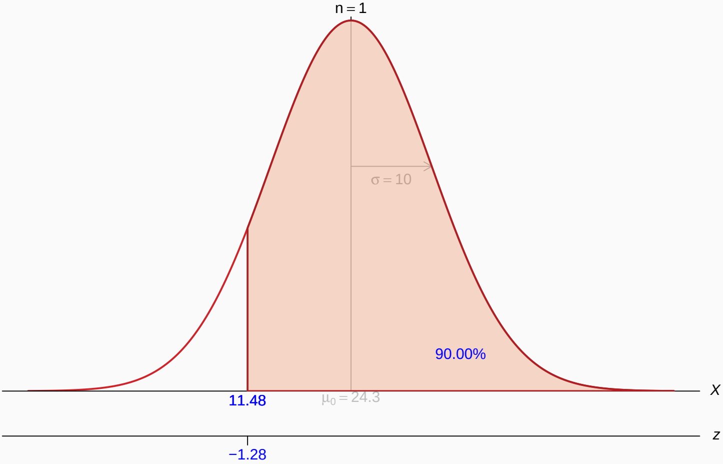
Find Scores Corresponding to Specific Proportions

- ```r
  qnorm(0.1, 24.3, 10, lower.tail = FALSE)
  ```
  ```
  ## [1] 37.11552
  ```
- ```r
  qnorm(1 - 0.1, 24.3, 10, lower.tail = TRUE)
  ```
  ```
  ## [1] 37.11552
  ```
- ```r
  q <- qnorm(0.1, 24.3, 10, lower.tail = TRUE)
  24.3 + (24.3 - q)
  ```
  ```
  ## [1] 37.11552
  ```
  ```

Find Scores Corresponding to Specific Proportions

- ```r
  qnorm(0.1, 24.3, 10, lower.tail = FALSE)
  ## [1] 37.11552
  ```
- ```r
  qnorm(1 - 0.1, 24.3, 10, lower.tail = TRUE)
  ## [1] 37.11552
  ```
- ```r
  q <- qnorm(0.1, 24.3, 10, lower.tail = TRUE)
  24.3 + (24.3 - q)
  ## [1] 37.11552
  ```
- ```r
  q <- qnorm(1 - 0.1, 24.3, 10, lower.tail = FALSE)
  24.3 + (24.3 - q)
  ## [1] 37.11552
  ```

Find Scores Corresponding to Specific Proportions



Find Scores Corresponding to Specific Proportions

Find Scores Corresponding to Specific Proportions

- Again, the distribution of commuting times for American workers is normal with a mean of $\mu = 24.3$ minutes and a standard deviation of $\sigma = 10$ minutes.

Find Scores Corresponding to Specific Proportions

- Again, the distribution of commuting times for American workers is normal with a mean of $\mu = 24.3$ minutes and a standard deviation of $\sigma = 10$ minutes.
- For this example, we will find the range of values that defines the middle 90% of the distribution.

Find Scores Corresponding to Specific Proportions

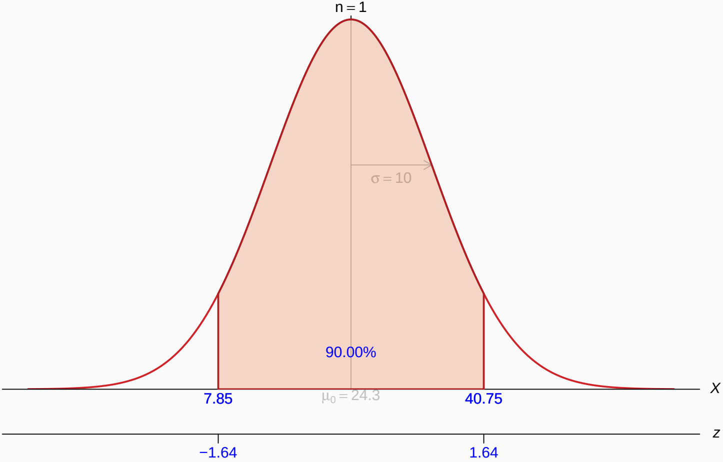
Find Scores Corresponding to Specific Proportions

- ```r
  qnorm(0.5 + 0.9 / 2, 24.3, 10, lower.tail = TRUE)
  ```
  ```
  ## [1] 40.74854
  ```
  ```r
  qnorm(0.5 - 0.9 / 2, 24.3, 10, lower.tail = TRUE)
  ```
  ```
  ## [1] 7.851464
  ```
  ```

Find Scores Corresponding to Specific Proportions

- ```
  qnorm(0.5 + 0.9 / 2, 24.3, 10, lower.tail = TRUE)
  ## [1] 40.74854
  qnorm(0.5 - 0.9 / 2, 24.3, 10, lower.tail = TRUE)
  ## [1] 7.851464
  ```
- ```
  qnorm((1 - 0.9) / 2, 24.3, 10, lower.tail = FALSE)
  ## [1] 40.74854
  qnorm((1 - 0.9) / 2, 24.3, 10, lower.tail = TRUE)
  ## [1] 7.851464
  ```

Find Scores Corresponding to Specific Proportions

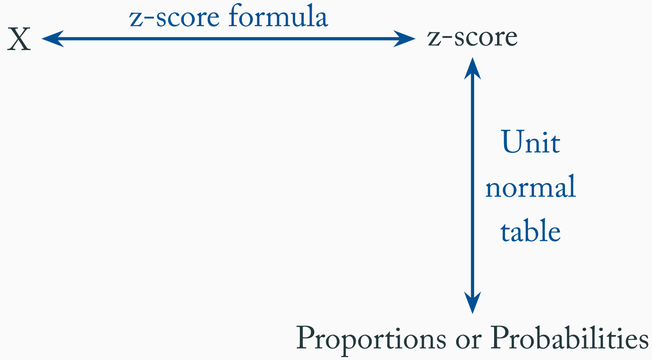


Relation between X, z-score, and proportions

X

Relation between X, z-score, and proportions

Relation between X, z-score, and proportions



Relation between X, z-score, and proportions

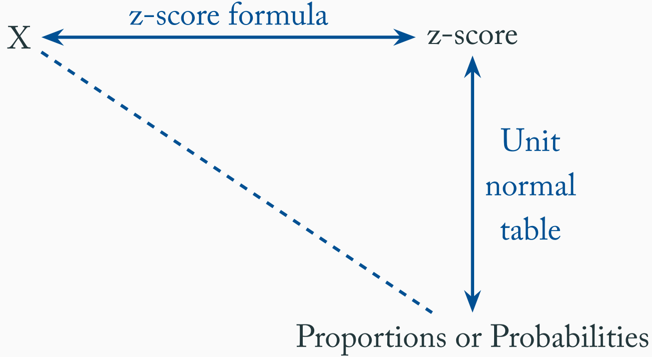


Table of Contents

The Binomial Distribution

Binomial and binomial data

Binomial and binomial data

- When a variable is measured on a scale consisting of exactly *two categories*, the resulting data are called binomial.

Binomial and binomial data

- When a variable is measured on a scale consisting of exactly *two categories*, the resulting data are called binomial.
- The term *binomial* can be loosely translated as "two names," referring to the two categories on the measurement scale.

Binomial and binomial data

Binomial and binomial data

- Binomial data can occur when a variable naturally exists with only two categories.

Binomial and binomial data

- Binomial data can occur when a variable naturally exists with only two categories.
- For example, people can be classified as male or female, and a coin toss results in either heads or tails.

Binomial and binomial data

- Binomial data can occur when a variable naturally exists with only two categories.
- For example, people can be classified as male or female, and a coin toss results in either heads or tails.
- It also is common for a researcher to simplify data by collapsing the scores into two categories.

Binomial and binomial data

- Binomial data can occur when a variable naturally exists with only two categories.
- For example, people can be classified as male or female, and a coin toss results in either heads or tails.
- It also is common for a researcher to simplify data by collapsing the scores into two categories.
- For example, a psychologist may use personality scores to classify people as either high or low in aggression.

Probability and the Binomial Distribution

Probability and the Binomial Distribution

- In binomial situations, the researcher often knows the probabilities associated with each of the two categories.

Probability and the Binomial Distribution

- In binomial situations, the researcher often knows the probabilities associated with each of the two categories.
- With a balanced coin, for example,

$$p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$$

Probability and the Binomial Distribution

- In binomial situations, the researcher often knows the probabilities associated with each of the two categories.
- With a balanced coin, for example,

$$p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$$

- The question of interest is the number of times each category occurs in a series of trials or in a sample of individuals. For example:

Probability and the Binomial Distribution

- In binomial situations, the researcher often knows the probabilities associated with each of the two categories.
- With a balanced coin, for example,

$$p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$$

- The question of interest is the number of times each category occurs in a series of trials or in a sample of individuals. For example:
- What is the probability of obtaining 15 heads in 20 tosses of a balanced coin?

Probability and the Binomial Distribution

- In binomial situations, the researcher often knows the probabilities associated with each of the two categories.
- With a balanced coin, for example,

$$p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$$

- The question of interest is the number of times each category occurs in a series of trials or in a sample of individuals. For example:
- What is the probability of obtaining 15 heads in 20 tosses of a balanced coin?
- What is the probability of obtaining more than 40 introverts in a sampling of 50 college freshmen?

The Binomial Distribution

The Binomial Distribution

- To answer probability questions about binomial data, we need to examine the binomial distribution.

The Binomial Distribution

- To answer probability questions about binomial data, we need to examine the binomial distribution.
- To define and describe this distribution, we first introduce some notation.

1. The two categories are identified as A and B.

The Binomial Distribution

- To answer probability questions about binomial data, we need to examine the binomial distribution.
- To define and describe this distribution, we first introduce some notation.

1. The two categories are identified as A and B.
2. The probabilities (or proportions) associated with each category are identified as

$$p = p(A) = \text{the probability of } A$$
$$q = p(B) = \text{the probability of } B$$

Notice that $p + q = 1.00$ because A and B are the only two possible outcomes.

The Binomial Distribution

The Binomial Distribution

3. The number of individuals or observations in the sample is identified by n.

The Binomial Distribution

3. The number of individuals or observations in the sample is identified by n.
4. The variable X refers to the number of times category A occurs in the sample.
   Notice that X can have any value from 0 (none of the sample is in category A) to n (all the sample is in category A).

- Using the notation presented here, the **binomial distribution** (二项式分布) shows the probability associated with each value of X from $X = 0$ to $X = n$.

The Binomial Distribution: An example

The Binomial Distribution: An example

- For this example, the event we are considering is a coin toss. There are two possible outcomes: heads and tails.

The Binomial Distribution: An example

- For this example, the event we are considering is a coin toss. There are two possible outcomes: heads and tails.
- We assume the coin is balanced, so

$$p = p(\text{heads}) = \frac{1}{2}$$
$$q = p(\text{tails}) = \frac{1}{2}$$

The Binomial Distribution: An example

- For this example, the event we are considering is a coin toss. There are two possible outcomes: heads and tails.
- We assume the coin is balanced, so

$$p = p(\text{heads}) = \frac{1}{2}$$
$$q = p(\text{tails}) = \frac{1}{2}$$

- We are looking at a sample of $n = 2$ tosses, and the variable of interest is

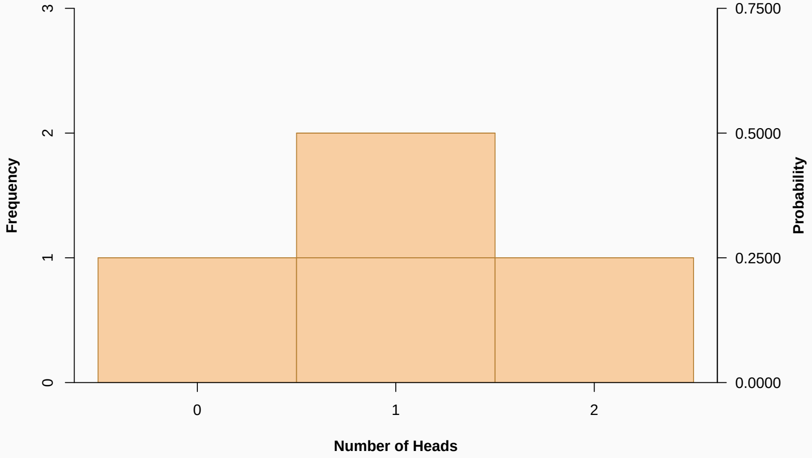
$$X = \text{the number of heads}$$

The Binomial Distribution: An example

- To construct the binomial distribution, we will look at all the possible outcomes from tossing a coin 2 times.
- The complete set of 4 outcomes is listed below.

	Toss-1	Toss-2	Number-of-Heads	Probability
1	Tail	Tail	0	0.2500
2	Head	Tail	1	0.2500
3	Tail	Head	1	0.2500
4	Head	Head	2	0.2500

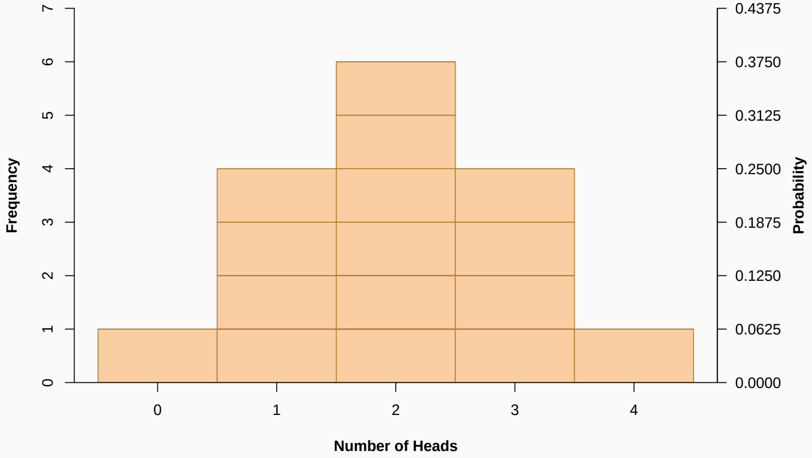
The Binomial Distribution: An example



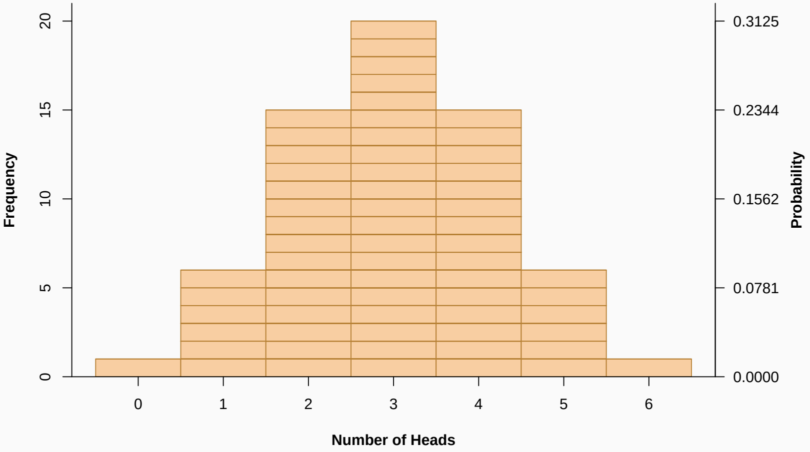
The Binomial Distribution: Another example

	Toss-1	Toss-2	Toss-3	Toss-4	Number-of-Heads	Probability
1	Tail	Tail	Tail	Tail	0	0.0625
2	Head	Tail	Tail	Tail	1	0.0625
3	Tail	Head	Tail	Tail	1	0.0625
4	Head	Head	Tail	Tail	2	0.0625
5	Tail	Tail	Head	Tail	1	0.0625
6	Head	Tail	Head	Tail	2	0.0625
7	Tail	Head	Head	Tail	2	0.0625
8	Head	Head	Head	Tail	3	0.0625
9	Tail	Tail	Tail	Head	1	0.0625
10	Head	Tail	Tail	Head	2	0.0625
11	Tail	Head	Tail	Head	2	0.0625
12	Head	Head	Tail	Head	3	0.0625
13	Tail	Tail	Head	Head	2	0.0625
14	Head	Tail	Head	Head	3	0.0625
15	Tail	Head	Head	Head	3	0.0625
16	Head	Head	Head	Head	4	0.0625

The Binomial Distribution: Another example



The Binomial Distribution: One more example



Binomial distribution functions

Function	Purpose	Example
dbinom(x, size, prob)	Probability Density Function (PDF)	*dbinom(1, 2, 0.5)* gives the density (height of the PDF) of the binomial 1 with size = 2 and prob = 0.5
pbinom(q, size, prob)	Cumulative Distribution Function (CDF)	*pbinom(1, 2, 0.5)* gives the area under the binomial curve to the left of 1, i.e., 0.75.
qbinom(p, size, prob)	Quantile Function - inverse of pbinom	*qbinom(0.75, 2, 0.5)* gives the value at which the CDF of the binomial is 0.75, i.e., 1.5
rbinom(n, size, prob)	Generate random numbers from binomial distribution	*rbinom(20, 2, 0.5)* generate 20 numbers from a binomial with size = 2 and probability = 0.5

Probability and the Binomial Distribution

Probability and the Binomial Distribution

- ```r
  pbinom(q = 0, size = 2,  prob = 0.5)
  ```
  ```
  ## [1] 0.25
  ```
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 2,  prob = 0.5)
  ## [1] 0.25
  ```
- ```
  pbinom(q = 1, size = 2,  prob = 0.5)
  ## [1] 0.75
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 2,  prob = 0.5)
  ## [1] 0.25
  ```

- ```
  pbinom(q = 1, size = 2,  prob = 0.5)
  ## [1] 0.75
  ```

- ```
  pbinom(q = 2, size = 2,  prob = 0.5)
  ## [1] 1
  ```
  ```

Probability and the Binomial Distribution

Probability and the Binomial Distribution

- ```r
  pbinom(q = 0, size = 4,  prob = 0.5)
  ```
  ```
  ## [1] 0.0625
  ```
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 4,  prob = 0.5)
  ## [1] 0.0625
  ```
- ```
  pbinom(q = 1, size = 4,  prob = 0.5)
  ## [1] 0.3125
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 4,  prob = 0.5)
  ## [1] 0.0625
  ```
- ```
  pbinom(q = 1, size = 4,  prob = 0.5)
  ## [1] 0.3125
  ```
- ```
  pbinom(q = 2, size = 4,  prob = 0.5)
  ## [1] 0.6875
  ```
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 4,  prob = 0.5)
  ## [1] 0.0625
  ```
- ```
  pbinom(q = 1, size = 4,  prob = 0.5)
  ## [1] 0.3125
  ```
- ```
  pbinom(q = 2, size = 4,  prob = 0.5)
  ## [1] 0.6875
  ```
- ```
  pbinom(q = 3, size = 4,  prob = 0.5)
  ## [1] 0.9375
  ```

Probability and the Binomial Distribution

- ```
  pbinom(q = 0, size = 4,  prob = 0.5)
  ## [1] 0.0625
  ```
- ```
  pbinom(q = 1, size = 4,  prob = 0.5)
  ## [1] 0.3125
  ```
- ```
  pbinom(q = 2, size = 4,  prob = 0.5)
  ## [1] 0.6875
  ```
- ```
  pbinom(q = 3, size = 4,  prob = 0.5)
  ## [1] 0.9375
  ```
- ```
  pbinom(q = 4, size = 4,  prob = 0.5)
  ## [1] 1
  ```
  ```

The Normal Approximation to the Binomial Distribution

The Normal Approximation

The Normal Approximation

- The binomial distribution tends to approximate a normal distribution, particularly when n is large.

The Normal Approximation

- The binomial distribution tends to approximate a normal distribution, particularly when n is large.
- To be more specific, the binomial distribution will be a nearly perfect normal distribution when pn and qn are both equal to or greater than 10.

The Normal Approximation

The Normal Approximation

- Under these circumstances, the binomial distribution will approximate a normal distribution with the following parameters:

$$\text{Mean: } \mu = pn \qquad \text{standard deviation: } \sigma = \sqrt{npq} \tag{1}$$

The Normal Approximation

- Under these circumstances, the binomial distribution will approximate a normal distribution with the following parameters:

$$\text{Mean: } \mu = pn \\ \text{standard deviation: } \sigma = \sqrt{npq} \tag{1}$$

- Within this normal distribution, each value of X has a corresponding z-score,

$$z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} \tag{2}$$

The Normal Approximation

- Under these circumstances, the binomial distribution will approximate a normal distribution with the following parameters:

$$\begin{aligned} \text{Mean: } \mu &= pn \\ \text{standard deviation: } \sigma &= \sqrt{npq} \end{aligned} \tag{1}$$

- Within this normal distribution, each value of X has a corresponding z-score,

$$z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} \tag{2}$$

- The fact that the binomial distribution tends to be normal in shape means that we can compute probability values directly from z-scores and the unit normal table.

The Normal Approximation

The Normal Approximation

- It is important to remember that the normal distribution is only an approximation of a true binomial distribution.

The Normal Approximation

- It is important to remember that the normal distribution is only an approximation of a true binomial distribution.
- Binomial values, such as the number of heads in a series of coin tosses, are *discrete*.

The Normal Approximation

- It is important to remember that the normal distribution is only an approximation of a true binomial distribution.
- Binomial values, such as the number of heads in a series of coin tosses, are *discrete*.
- For example, if you are recording the number of heads in four tosses of a coin, you may observe 2 heads or 3 heads but there are no values between 2 and 3.

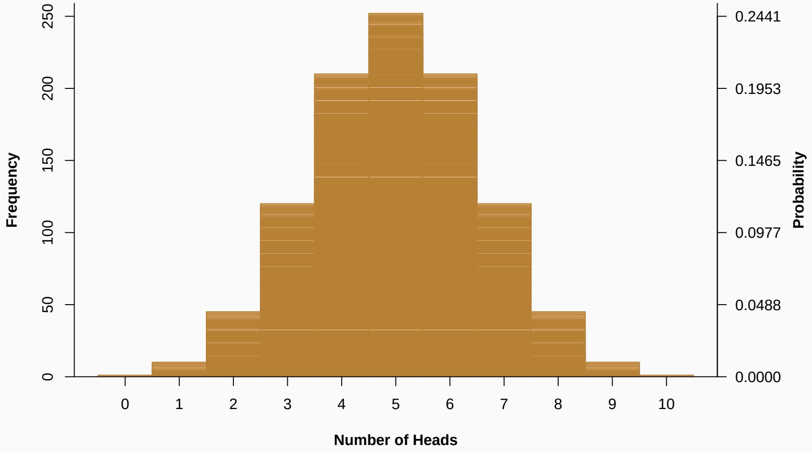
The Normal Approximation

- It is important to remember that the normal distribution is only an approximation of a true binomial distribution.
- Binomial values, such as the number of heads in a series of coin tosses, are *discrete*.
- For example, if you are recording the number of heads in four tosses of a coin, you may observe 2 heads or 3 heads but there are no values between 2 and 3.
- The normal distribution, on the other hand, is *continuous*.

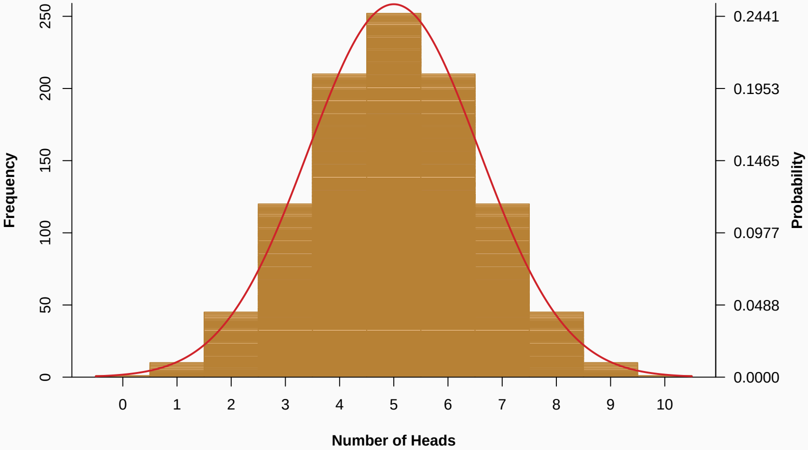
The Normal Approximation

- It is important to remember that the normal distribution is only an approximation of a true binomial distribution.
- Binomial values, such as the number of heads in a series of coin tosses, are *discrete*.
- For example, if you are recording the number of heads in four tosses of a coin, you may observe 2 heads or 3 heads but there are no values between 2 and 3.
- The normal distribution, on the other hand, is *continuous*.
- However, the normal approximation provides an extremely accurate model for computing binomial probabilities in many situations.

Probability and the Binomial Distribution



Probability and the Binomial Distribution



The Normal Approximation

The Normal Approximation

- To gain maximum accuracy when using the normal approximation, you must remember that each X value in the binomial distribution actually corresponds to a bar in the histogram.

The Normal Approximation

- To gain maximum accuracy when using the normal approximation, you must remember that each X value in the binomial distribution actually corresponds to a bar in the histogram.
- For example, the score $X = 6$ in a histogram is represented by a bar that is bounded by real limits of 5.5 and 6.5.

The Normal Approximation

- To gain maximum accuracy when using the normal approximation, you must remember that each X value in the binomial distribution actually corresponds to a bar in the histogram.
- For example, the score $X = 6$ in a histogram is represented by a bar that is bounded by real limits of 5.5 and 6.5.
- The actual probability of X = 6 is determined by the area contained in this bar.

The Normal Approximation

The Normal Approximation

- To approximate this probability using the normal distribution, you should find the area that is contained between the two real limits.

The Normal Approximation

- To approximate this probability using the normal distribution, you should find the area that is contained between the two real limits.
- Similarly, if you are using the normal approximation to find the probability of obtaining a score greater than $X = 6$, you should use the area beyond the real limit boundary of 6.5.

The Normal Approximation

The Normal Approximation

- Suppose that you plan to test for ESP (extra-sensory perception) by asking people to predict the suit of a card that is randomly selected from a complete deck.

The Normal Approximation

- Suppose that you plan to test for ESP (extra-sensory perception) by asking people to predict the suit of a card that is randomly selected from a complete deck.
- What is the probability for an individual without ESP to guess correctly more than 15 out of 48 trials.

The Normal Approximation

- Suppose that you plan to test for ESP (extra-sensory perception) by asking people to predict the suit of a card that is randomly selected from a complete deck.
- What is the probability for an individual without ESP to guess correctly more than 15 out of 48 trials.
- ```r
  (bp <- pbinom(15, 48, 1/4, lower.tail = FALSE))
  ## [1] 0.1231778
  ```
  ```

Probability and the Binomial Distribution

```
n <- 48
p <- 1/4
q <- 1 - p
M <- p * n
S <- sqrt(n * p * q)
```

Probability and the Binomial Distribution

```r
n <- 48
p <- 1/4
q <- 1 - p
M <- p * n
S <- sqrt(n * p * q)
```

```r
(z <- (15.5 - M) / S)

## [1] 1.166667
```

Probability and the Binomial Distribution

```
n <- 48
p <- 1/4
q <- 1 - p
M <- p * n
S <- sqrt(n * p * q)
```

```
(z <- (15.5 - M) / S)

## [1] 1.166667
```

```
(np1 <- pnorm(z, lower.tail = FALSE))

## [1] 0.1216725
```

Probability and the Binomial Distribution

```
bp - pnorm(15.5, M, S, lower.tail = FALSE)

## [1] 0.001505259
```

Probability and the Binomial Distribution

```r
bp - pnorm(15.5, M, S, lower.tail = FALSE)

## [1] 0.001505259
```

```r
bp - pnorm(15, M, S, lower.tail = FALSE)

## [1] -0.03547749
```

Probability and the Binomial Distribution

```r
bp - pnorm(15.5, M, S, lower.tail = FALSE)

## [1] 0.001505259
```

```r
bp - pnorm(15, M, S, lower.tail = FALSE)

## [1] -0.03547749
```

```r
bp - pnorm(16, M, S, lower.tail = FALSE)

## [1] 0.03196654
```

Probability and the Binomial Distribution

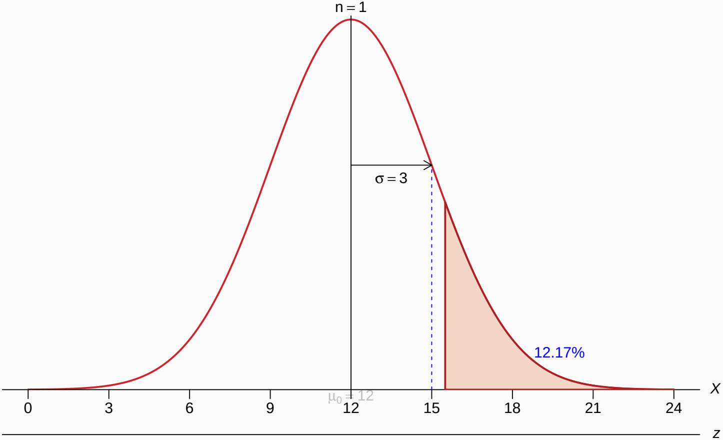


Table of Contents

Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- Probability forms a direct link between samples and the populations from which they come.

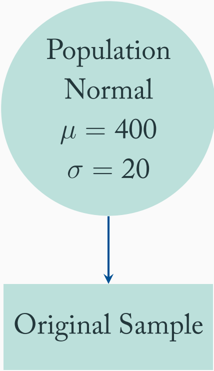
Looking Ahead to Inferential Statistics

- Probability forms a direct link between samples and the populations from which they come.
- This link is the foundation for the inferential statistics in future chapters.

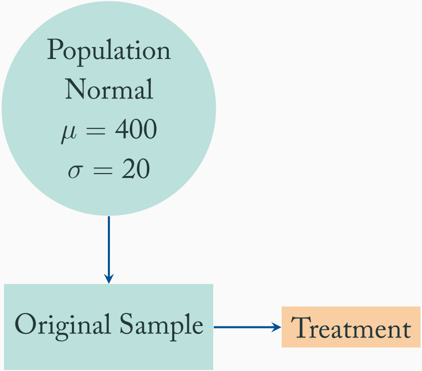
Looking Ahead to Inferential Statistics

Population
Normal
$\mu = 400$
$\sigma = 20$

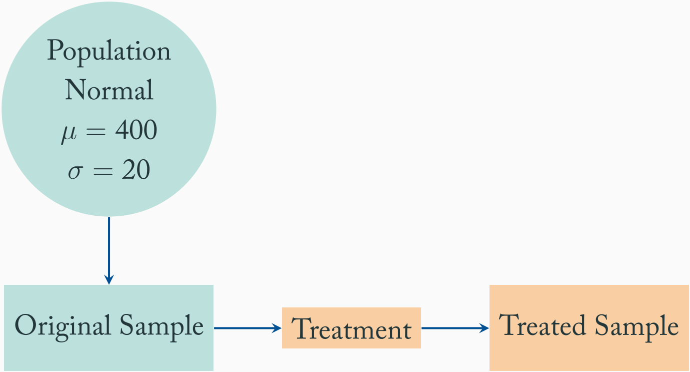
Looking Ahead to Inferential Statistics



Looking Ahead to Inferential Statistics



Looking Ahead to Inferential Statistics



Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- To determine whether the treatment has an effect, the researcher simply compares the treated sample with the original population.

Looking Ahead to Inferential Statistics

- To determine whether the treatment has an effect, the researcher simply compares the treated sample with the original population.
- If the individuals in the sample have scores around 400 (the original population mean), then we must conclude that the treatment appears to have no effect.

Looking Ahead to Inferential Statistics

- To determine whether the treatment has an effect, the researcher simply compares the treated sample with the original population.
- If the individuals in the sample have scores around 400 (the original population mean), then we must conclude that the treatment appears to have no effect.
- On the other hand, if the treated individuals have scores that are noticeably different from 400, then the researcher has evidence that the treatment does have an effect.

Looking Ahead to Inferential Statistics

- To determine whether the treatment has an effect, the researcher simply compares the treated sample with the original population.
- If the individuals in the sample have scores around 400 (the original population mean), then we must conclude that the treatment appears to have no effect.
- On the other hand, if the treated individuals have scores that are noticeably different from 400, then the researcher has evidence that the treatment does have an effect.
- Notice that the study is using a sample to help answer a question about a population; this is the essence of inferential statistics.

Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- The problem for the researcher is determining exactly what is meant by "noticeably different" from 400.

Looking Ahead to Inferential Statistics

- The problem for the researcher is determining exactly what is meant by "noticeably different" from 400.
- In Chapter 5, we suggested that z-scores provide one method for solving this problem. Specifically, we suggested that a z-score value beyond $z = 2.00$ (or -2.00) was an extreme value and therefore noticeably different. However, the choice of $z = \pm 2.00$ was purely arbitrary.

Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- Now we have another tool, *probability*, to help us decide exactly where to set the boundaries.

Looking Ahead to Inferential Statistics

- Now we have another tool, *probability*, to help us decide exactly where to set the boundaries.
- Middle 95% High probability values (scores near m 5 400) indicating that the treatment has no effect.

Looking Ahead to Inferential Statistics

- Now we have another tool, *probability*, to help us decide exactly where to set the boundaries.
- Middle 95% High probability values (scores near m 5 400) indicating that the treatment has no effect.
- Extreme 5% Scores that are very unlikely to be obtained from the original population and therefore provide evidence of a treatment effect.

Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- Using column C of the unit normal table, the z-score boundaries for the right and left tails are $z = +1.96$ and $z = -1.96$, respectively.

Looking Ahead to Inferential Statistics

- Using column C of the unit normal table, the z-score boundaries for the right and left tails are $z = +1.96$ and $z = -1.96$, respectively.
- If we are selecting an individual from the original untreated population, then it is very unlikely that we would obtain a score beyond the $z = \pm 1.96$ boundaries.

Looking Ahead to Inferential Statistics

- Using column C of the unit normal table, the z-score boundaries for the right and left tails are $z = +1.96$ and $z = -1.96$, respectively.
- If we are selecting an individual from the original untreated population, then it is very unlikely that we would obtain a score beyond the $z = \pm 1.96$ boundaries.
- The boundaries set at $z = \pm 1.96$ provide objective criteria for deciding whether our sample provides evidence that the treatment has an effect.

Looking Ahead to Inferential Statistics

Looking Ahead to Inferential Statistics

- Specifically, if our sample is located in the tail beyond one of the $z = \pm 1.96$ boundaries, then we can conclude:

Looking Ahead to Inferential Statistics

- Specifically, if our sample is located in the tail beyond one of the $z = \pm 1.96$ boundaries, then we can conclude:
    1. The sample is an extreme value, nearly 2 standard deviations away from average, and therefore is noticeably different from most individuals in the original population.

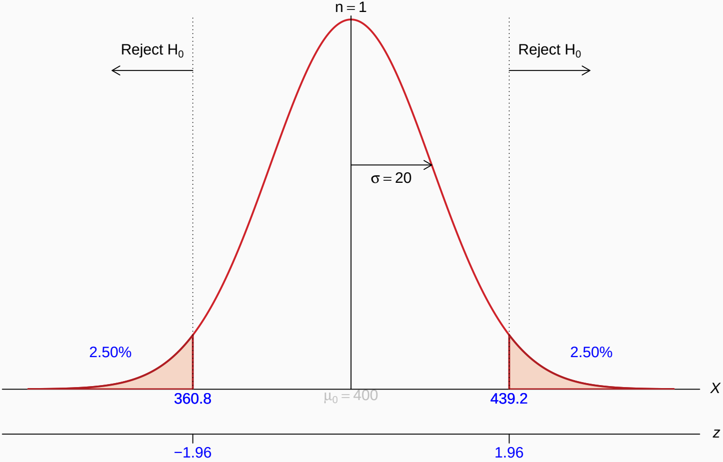
Looking Ahead to Inferential Statistics

- Specifically, if our sample is located in the tail beyond one of the $z = \pm 1.96$ boundaries, then we can conclude:
  1. The sample is an extreme value, nearly 2 standard deviations away from average, and therefore is noticeably different from most individuals in the original population.
  2. The sample is a very unlikely value with a very low probability if the treatment has no effect.

Looking Ahead to Inferential Statistics

- Specifically, if our sample is located in the tail beyond one of the $z = \pm 1.96$ boundaries, then we can conclude:
    1. The sample is an extreme value, nearly 2 standard deviations away from average, and therefore is noticeably different from most individuals in the original population.
    2. The sample is a very unlikely value with a very low probability if the treatment has no effect.
- Therefore, the sample provides clear evidence that the treatment has had an effect.

Looking Ahead to Inferential Statistics



Questions?