

Lesson 6: Normal distribution

TB sections 3.3

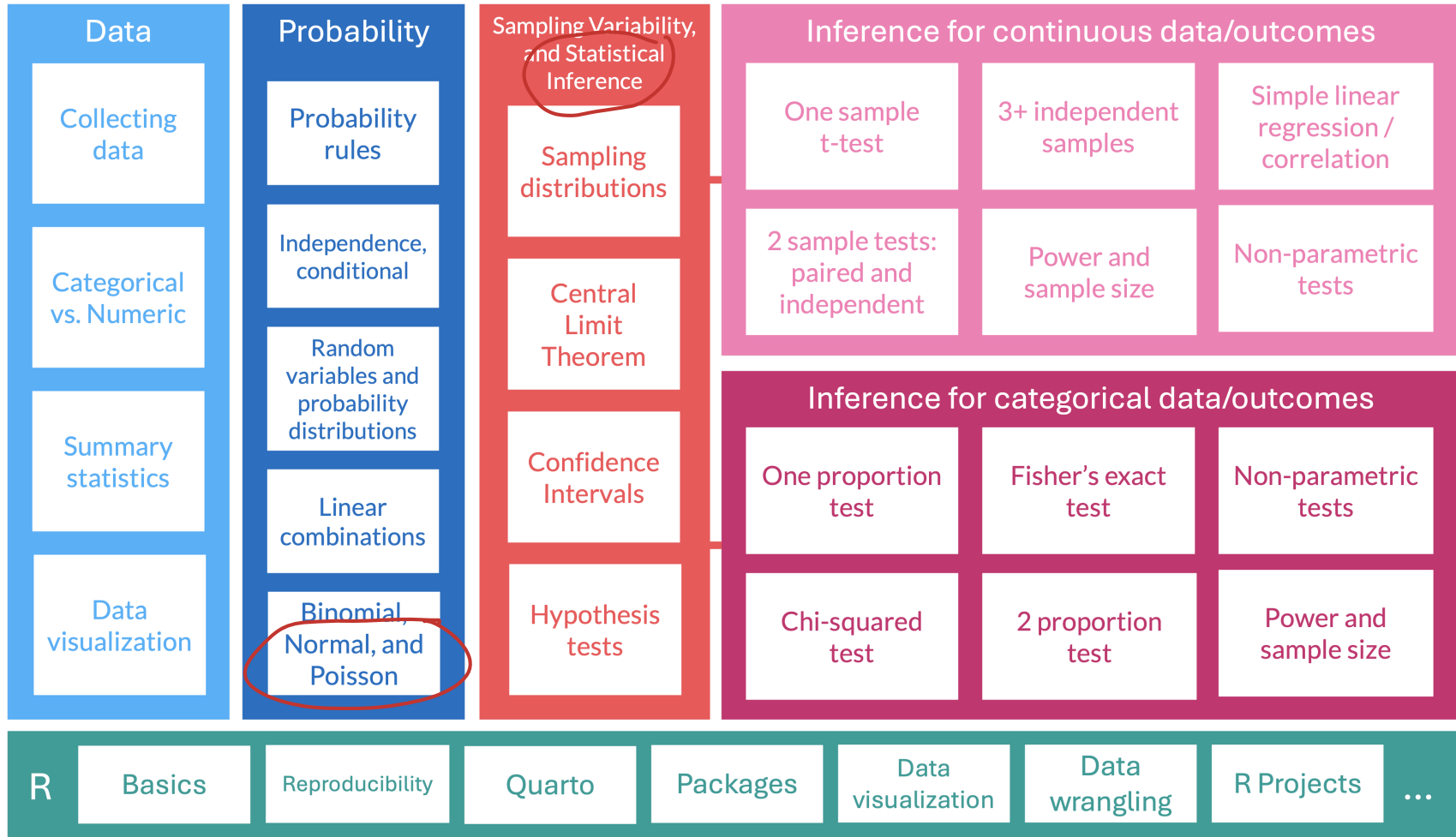
Meike Niederhausen and Nicky Wakim

2025-10-20

Learning Objectives

1. Understand how probability distributions extend to continuous distributions
2. Calculate probabilities for specific events using a Normal distribution
3. Apply the Normal distribution to approximate probabilities for binomial events
discrete

Where are we?



Learning Objectives

1. Understand how probability distributions extend to continuous distributions
2. Calculate probabilities for specific events using a Normal distribution
3. Apply the Normal distribution to approximate probabilities for binomial events

Last time: Discrete vs. continuous random variables

- Probability distributions are usually either **discrete** or **continuous**, depending on whether the random variable is discrete or continuous.

Discrete random variable

A **discrete** r.v. X takes on a finite number of values or countably infinite number of possible values.

Think:

- Number of heads in a set of coin tosses
- Number of people who have had chicken pox in a random sample

- Binomial and Bernoulli distributions are discrete

Continuous random variable

A **continuous** r.v. X can take on any real value in an interval of values or unions of intervals.

Think:

- Height in a population
- Blood pressure in a population

1.5
1.49
1.499

Probabilities for continuous distributions (1/2)

$$P(X=x) = 0$$

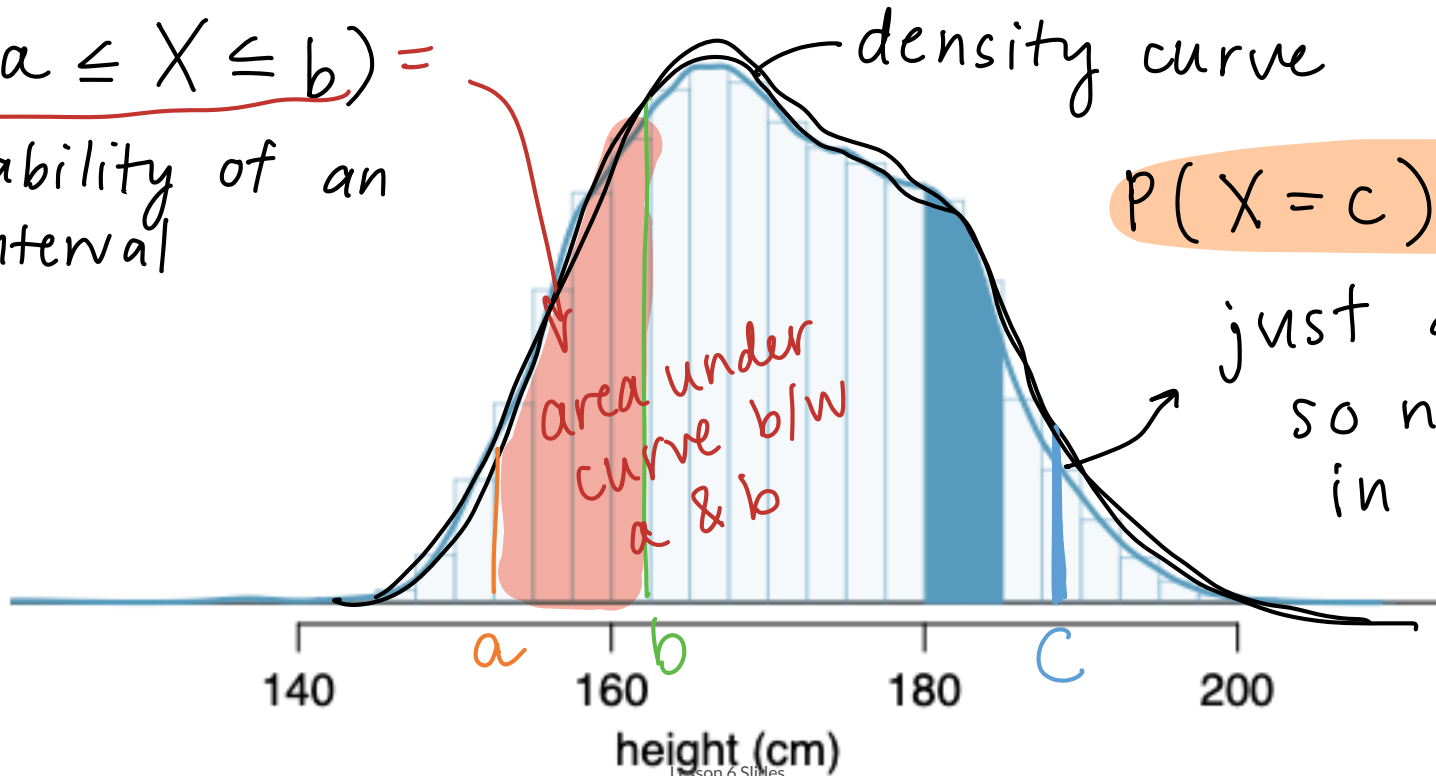
in cont distributions

Two important features of continuous distributions:

- The total area under the density curve is 1.
- The probability that a variable has a value within a specified interval is the area under the curve over that interval.

$$\hookrightarrow P(a \leq X \leq b) =$$

probability of an interval



$$P(X=c) = 0$$

Probabilities for continuous distributions (2/2)

When working with continuous random variables, probability is found for intervals of values rather than individual values.

- The probability that a continuous r.v. X takes on any single individual value is 0
 - That is, $P(X = x) = 0$.
- Thus, $P(a < X < b)$ is equivalent to $P(a \leq X \leq b)$

$$P(a < X < b) = P(a \leq X \leq b)$$

$$0 \leq P(a < X < b) \leq 1$$

Poll Everywhere Question 1

13:26 Mon Oct 20

86%

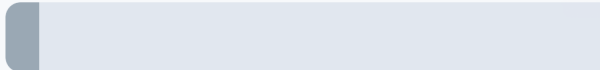


Join by Web PollEv.com/nickywakim275



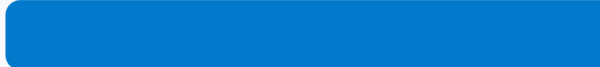
Which of the following statements is true about the probability distribution of a continuous random variable?

The probability that the random variable takes any specific value is greater than zero



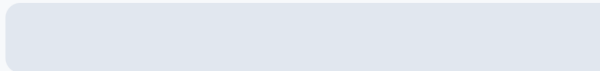
5%

The probability that the random variable takes any specific value is exactly zero



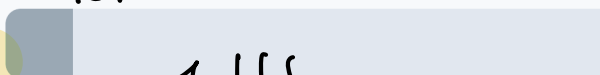
86%

The probability of any interval of values is always zero



0%

The total area under the ~~probability density function~~ probability density function (PDF) curve is less than 1



10%

$P(X=x) > 0$ X

$P(X=x) = 0$ ✓

$P(a \leq X \leq b) = 0$ X

$\int f(x) < 1$ X

distribution
is 1!!!

Learning Objectives

1. Understand how probability distributions extend to continuous distributions
2. Calculate probabilities for specific events using a Normal distribution
3. Apply the Normal distribution to approximate probabilities for binomial events

Normal distribution

- A random variable X is modeled with a normal distribution if:

- **Shape:** symmetric, unimodal bell curve
- **Center:** mean μ \hookrightarrow one hump
- **Spread (variability):** standard deviation σ or variance σ^2

- Shorthand for a random variable, X , that has a Normal distribution:

$$X \sim \text{Normal}(\mu, \sigma)$$

\hookrightarrow "distributed as"

- **Example:** We recorded the high temperature in the past 100 years for today. The mean high is 19°C (66.2°F)

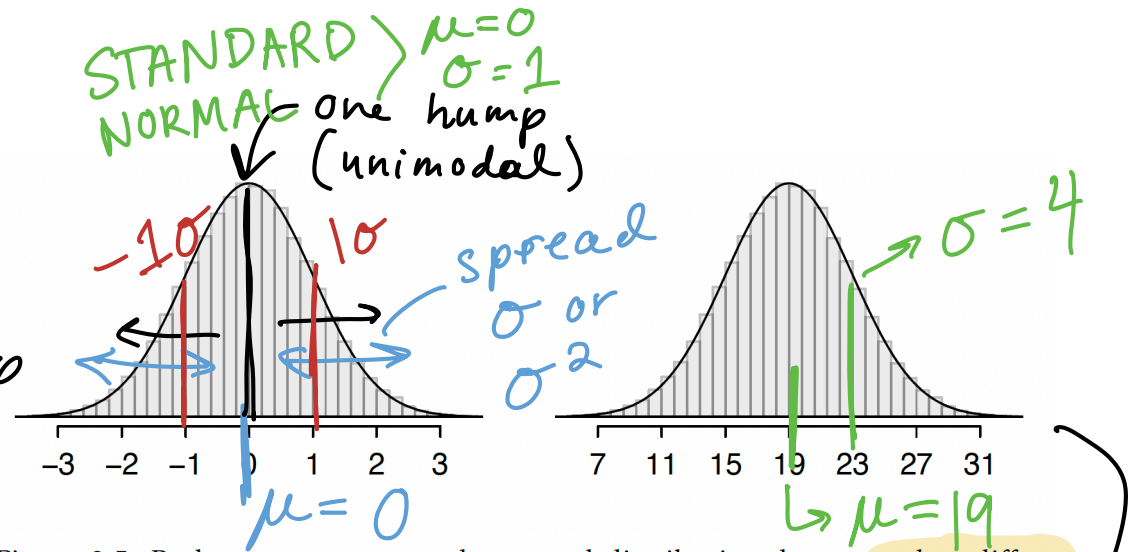


Figure 3.5: Both curves represent the normal distribution; however, they differ in their center and spread. The normal distribution with mean 0 and standard deviation 1 is called the **standard normal distribution**.

$$X \sim \text{Normal}(19, 4)$$

Standard Normal distribution (1/2)

$$\sigma^2 = \frac{1}{1} = 1$$
$$\sigma = \sqrt{1} = 1$$

- A *standard normal* distribution is defined as a normal distribution with mean 0 and variance 1. It is often denoted as $Z \sim N(0, 1)$.
- Any normal random variable X can be transformed into a standard normal random variable Z .

↪ standardization

$$X \sim N(\mu, \sigma)$$

$$Z = \frac{X - \mu}{\sigma}$$

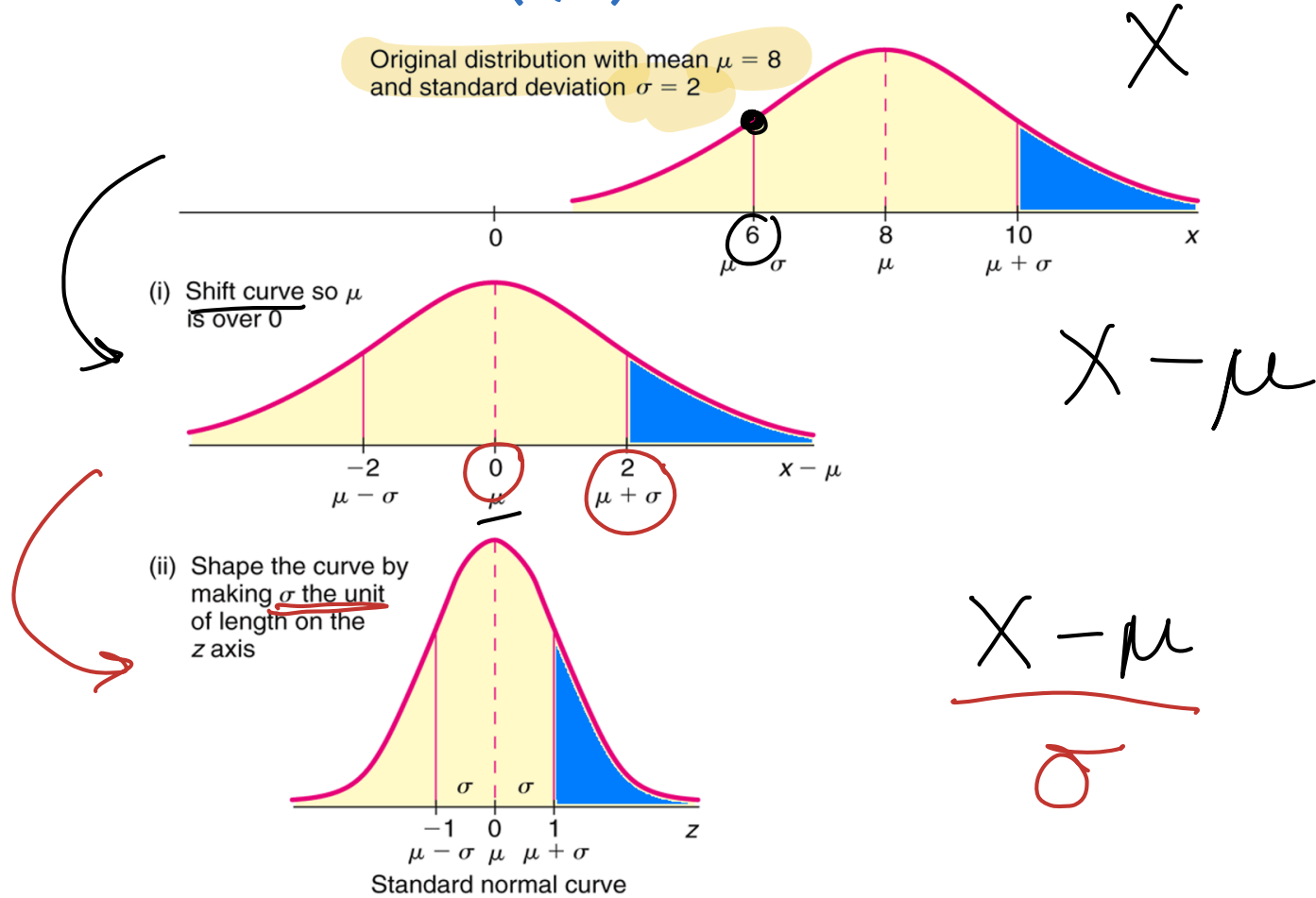
$$X = \mu + Z\sigma$$

$$\frac{X - \mu}{\sigma} \sim N(0, 1)$$

- The Z -score of an observation quantifies how far the observation is from the mean, in units of standard deviation(s).
- For example, if an observation has Z -score $z = 3.4$, then the observation is 3.4 standard deviations above the mean.

Z -score of -3.4 below

Standard Normal distribution (2/2)



Transformation from general normal X to standard normal Z

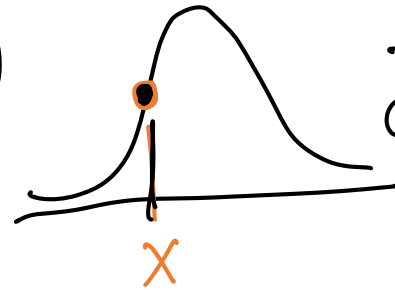
Normal distribution: R commands

R commands with their **input** and **output**:

R code	What does it return?
<code>rnorm()</code>	returns <u>sample</u> of random variables with specified normal distribution
<code>dnorm()</code>	returns value of probability density at certain point of the normal distribution <ul style="list-style-type: none">• Not typically used bc this is a continuous distributon
<code>pnorm()</code>	returns <u>cumulative probability</u> of getting certain point (or less) of the normal distribution
<code>qnorm()</code>	returns <u>z-score</u> corresponding to desired quantile

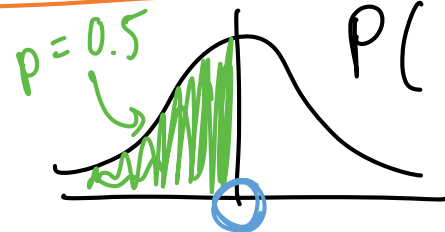
sample of 100 based on dist.

$f(x)$



$$\frac{P(X=x)=0}{dnorm \neq P(X=x)}$$

$P(X \leq x)$



output input

$$P(X \leq \overset{\text{output}}{\underset{\text{input}}{x}}) = 0.5$$

what is x?

x -value

Calculating probabilities from a Normal distribution

Three ways to calculate probabilities from a normal distribution:

1. Calculus (not for us!)

2. Normal probability table

- The textbook has a normal probability table in Appendix B.1, which is included as the next two pages
- Not required for this class

3. R commands

• $P(Z \leq q) = \text{pnorm}(q, \text{mean} = 0, \text{sd} = 1, \text{lower.tail} = \text{TRUE})$

• $P(Z > q) = \text{pnorm}(q, \text{mean} = 0, \text{sd} = 1, \text{lower.tail} = \text{FALSE})$

$\hookrightarrow P(Z \geq q)$

$\hookrightarrow P(X > q)$

$P(X \leq q)$

4. ~~Random~~ online calculators

- Like this one: https://onlinestatbook.com/2/calculators/normal_dist.html

Example: Calculating probabilities from a Normal distribution (1/5)

Example: Calculating standard normal probabilities practice

Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$. Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

1. $\mathbb{P}(Z < 2.67)$
2. $\mathbb{P}(Z > -0.37)$
3. $\mathbb{P}(-2.18 < Z < 2.46)$
4. $\mathbb{P}(Z = 1.53)$

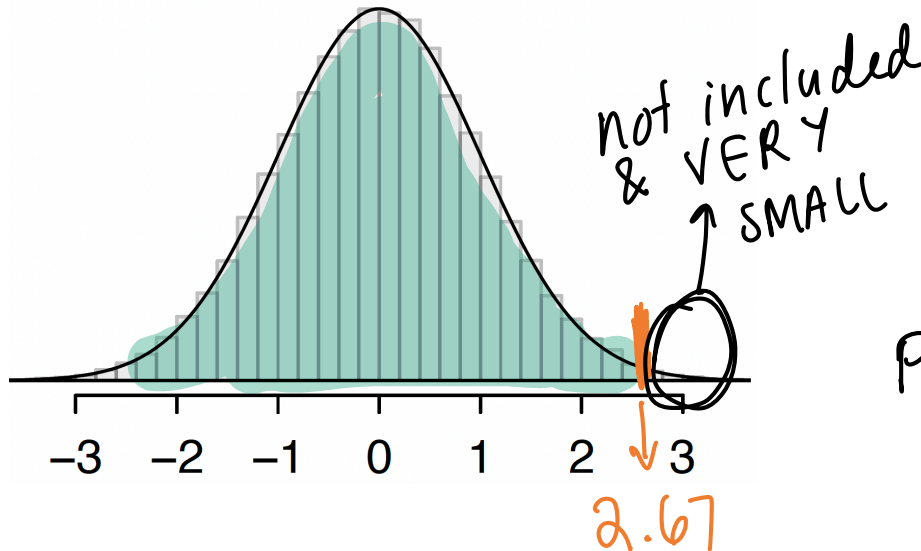
Example: Calculating probabilities from a Normal distribution (2/5)

Example: Calculating standard normal probabilities practice

Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$. Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

1. $\mathbb{P}(Z < 2.67)$

1. Draw on standard Normal curve:



$$\mathbb{P}(Z < 2.67)$$

2. Calculate probability:

```
1 pnorm(q = 2.67, mean = 0, sd = 1)
```

```
[1] 0.9962074
```

```
1 pnorm(q = 2.67)
```

default values

```
[1] 0.9962074
```

$$\mathbb{P}(Z < 2.67) = 0.996$$

Example: Calculating probabilities from a Normal distribution (3/5)

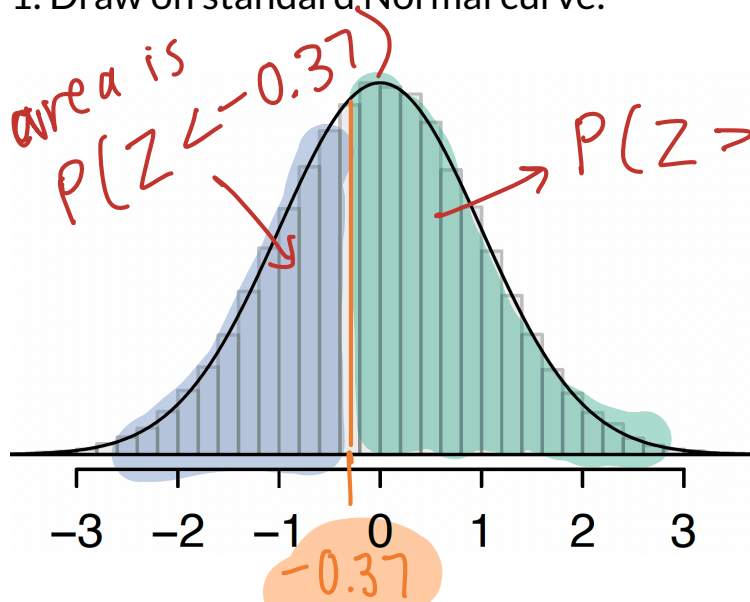
Example: Calculating standard normal probabilities practice

Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$. Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

2. $\mathbb{P}(Z > -0.37)$

$$P(Z > -0.37) = 1 - P(Z < -0.37)$$

1. Draw on standard Normal curve:



```
1 1 - pnorm(q = -0.37,  
2 mean = 0,  
3 sd = 1)
```

```
[1] 0.6443088
```

```
1 pnorm(q = -0.37, mean = 0, sd = 1,  
2 lower.tail = FALSE)
```

```
[1] 0.6443088
```

default is
 $\text{lower.tail} = T$

$\hookrightarrow P(Z > -0.37)$

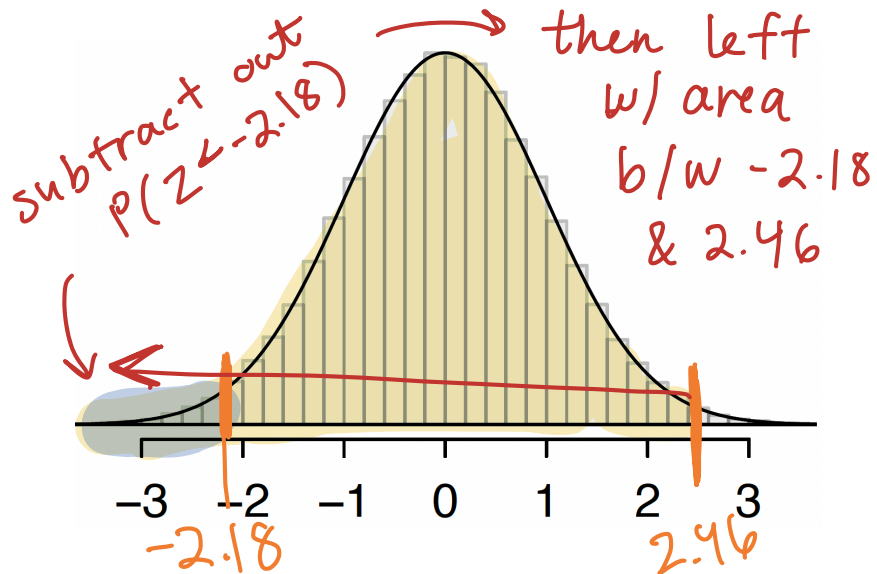
Example: Calculating probabilities from a Normal distribution (4/5)

Example: Calculating standard normal probabilities practice

Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$. Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

$$3. \mathbb{P}(-2.18 < Z < 2.46) = P(Z < 2.46) - P(Z < -2.18)$$

1. Draw on standard Normal curve:



```
1 pnorm(q = 2.46, mean = 0, sd = 1) -  
2 pnorm(q = -2.18, mean = 0, sd = 1)  
[1] 0.9784244
```

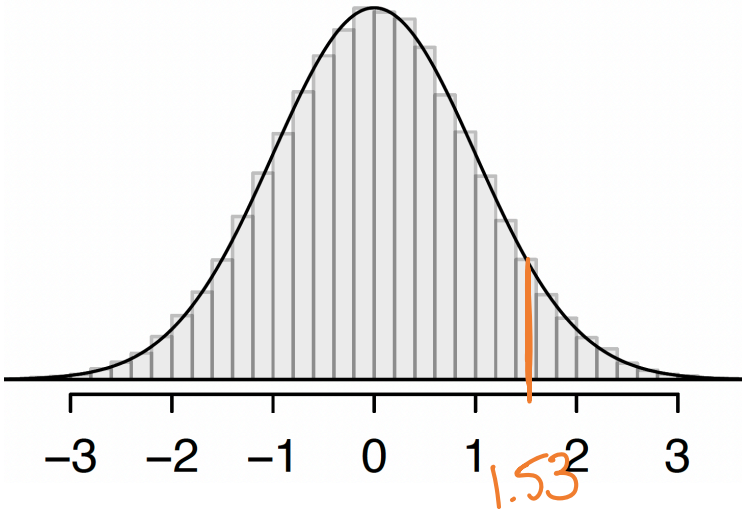
Example: Calculating probabilities from a Normal distribution (5/5)

Example: Calculating standard normal probabilities practice

Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$. Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

4. $\mathbb{P}(Z = 1.53)$

1. Draw on standard Normal curve:



$P(X=x) = 0$ for cont.
distributions

$$P(Z = 1.53) = 0$$

Example: Using Normal distribution in word problems (1/4)

Example: Diastolic blood pressure (DBP)

Suppose the distribution of diastolic blood pressure (DBP) in 35- to 44-year old men is normally distributed with mean 80 mm Hg and variance 144 mm Hg.

1. Mild hypertension is when the DBP is between 90 and 99 mm Hg. What proportion of this population has mild hypertension?
2. What is the 10th percentile of the DBP distribution?
3. What is the 95th percentile of the DBP distribution?

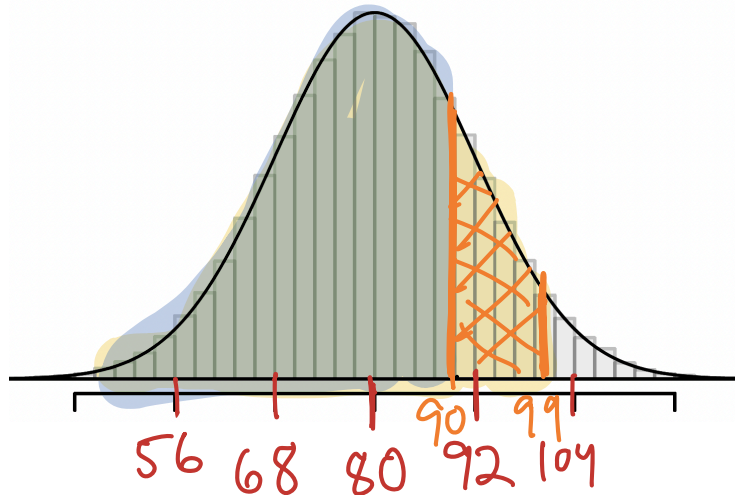
Example: Using Normal distribution in word problems (2/4)

Example: Diastolic blood pressure (DBP)

Suppose the distribution of diastolic blood pressure (DBP) in 35- to 44-year old men is normally distributed with mean 80 mm Hg and variance 144 mm²Hg. $sd = \sqrt{var} = \sqrt{144} = 12$

1. Mild hypertension is when the DBP is between 90 and 99 mm Hg. What proportion of this population has mild hypertension?

- Draw on a normal curve:



- Compute in R:

```
1 pnorm(q = 99, mean = 80,  
2      sd = sqrt(144)) -  
3 pnorm(q = 90, mean = 80,  
4      sd = sqrt(144))
```

```
[1] 0.1456556
```

$$\downarrow$$
$$\text{sqrt}(x) = \sqrt{x}$$

Example: Using Normal distribution in word problems (3/4)

Example: Diastolic blood pressure (DBP)

Suppose the distribution of diastolic blood pressure (DBP) in 35- to 44-year old men is normally distributed with mean 80 mm Hg and variance 144 mm Hg.

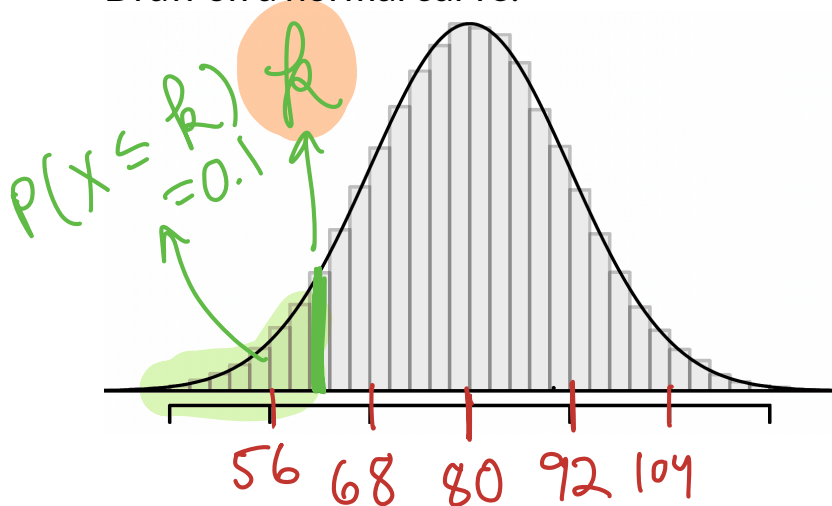
2. What is the 10th percentile of the DBP distribution?

find k so that

$$P(X \leq k) = 0.1$$

$P = 0.1$

- Draw on a normal curve:



- Compute in R:

```
1 qnorm(p = 0.10,  
2       mean = 80,  
3       sd = sqrt(144))
```

```
[1] 64.62138
```

$$P(X \leq 64.621) = 0.1$$

10th percentile is 64.621 mmHg

Example: Using Normal distribution in word problems (4/4)

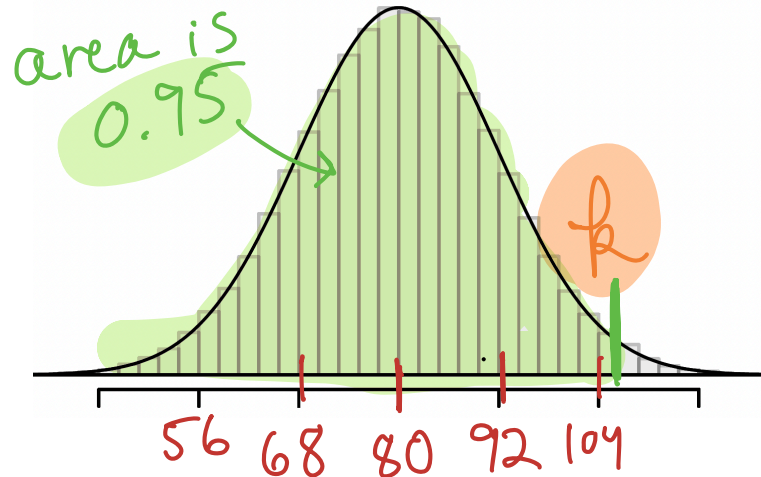
Example: Diastolic blood pressure (DBP)

Suppose the distribution of diastolic blood pressure (DBP) in 35- to 44-year old men is normally distributed with mean 80 mm Hg and variance 144 mm Hg.

3. What is the 95th percentile of the DBP distribution?

$$P(X \leq k) = 0.95$$

- Draw on a normal curve:



- Compute in R:

```
1 qnorm(p = 0.95,  
2       mean = 80,  
3       sd = sqrt(144))
```

```
[1] 99.73824
```

$$P(X \leq 99.738) = 0.95$$

95th percentile is 99.738 mm Hg

Learning Objectives

1. Understand how probability distributions extend to continuous distributions
2. Calculate probabilities for specific events using a Normal distribution
3. Apply the Normal distribution to approximate probabilities for binomial events

Normal Approximation of the Binomial Distribution

- Recall that a binomial random variable X counts the total number of successes in n independent trials, each with probability p of a success.

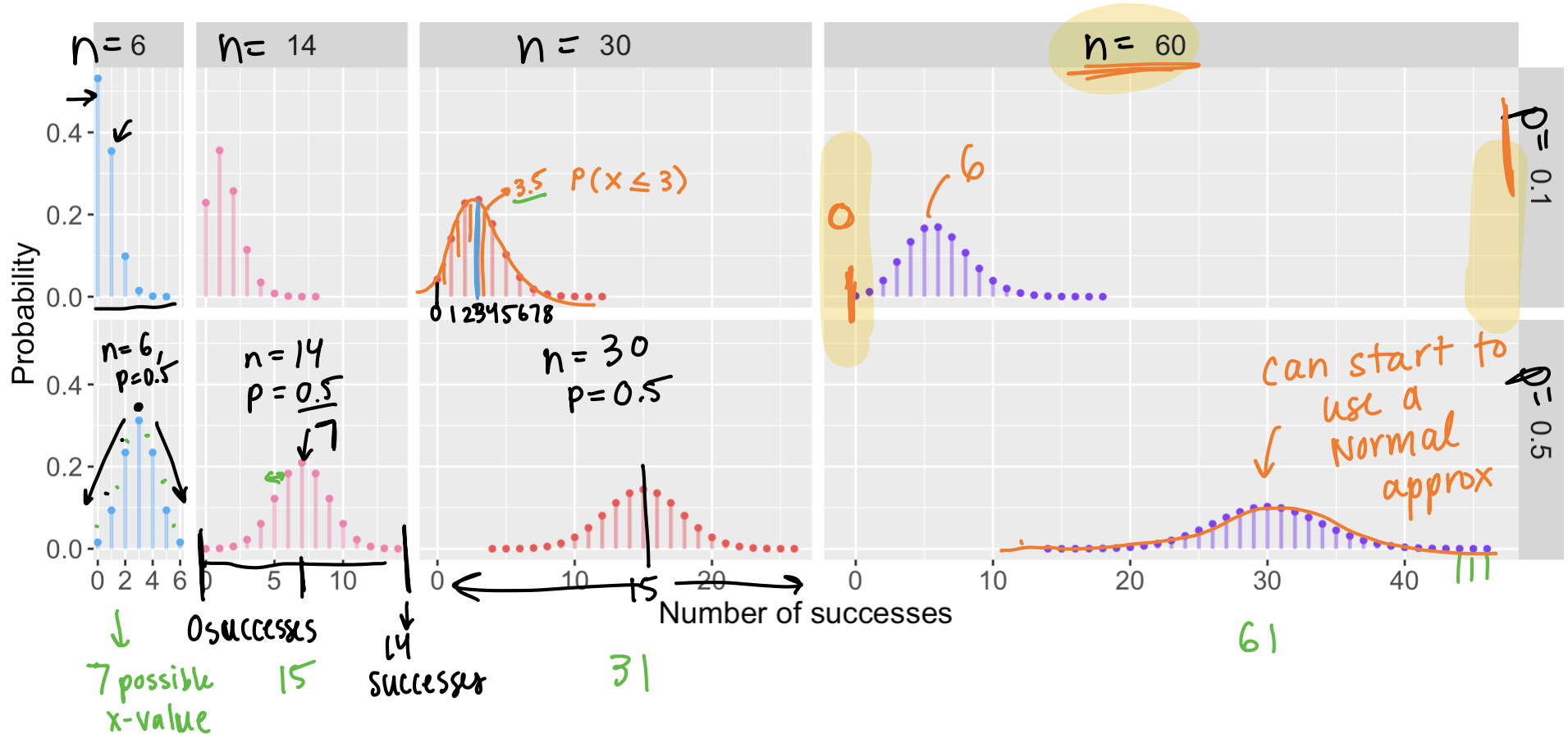
- Probability function for $k = 0, 1, \dots, n$:

$$\underline{P(X = k)} = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} \quad \left. \vphantom{\frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k}} \right\} \text{ discrete}$$

- Tedious to compute for large number of trials (n), although doable with software like R
- As n gets big though, the distribution shape of a binomial r.v. gets more and more symmetric, and can be approximated by a normal distribution
- Pretty good video behind the intuition of this (Watch 00:00 - 05:40)

We can look at a plot of Binomial distributions

- Binomial distributions for different n (columns) and p (rows)



Normal Approximation of the Binomial Distribution

- Also known as: **Sampling distribution of \hat{p}**
- If $X \sim \text{Binomial}(n, p)$ and $np > 10$ and $nq = n(1 - p) > 10$
 - Ensures sample size (n) is moderately large and the p is not too close to 0 or 1
 - Other resources use other criteria (like $npq > 5$ or $np > 5$)

$$\hookrightarrow np(1-p) > 5$$

- THEN approximately

$$X \sim \text{Normal}(\mu_X = np, \sigma_X = \sqrt{np(1-p)})$$

- **Continuity Correction:** Applied to account for the fact that the binomial distribution is discrete, while the normal distribution is continuous
 - Adjust the binomial value (# of successes) by ± 0.5 before calculating the normal probability.
 - For $P(X \leq k)$ (Binomial), you would instead calculate $P(X \leq k + 0.5)$ (Normal approx)
 - For $P(X \geq k)$ (Binomial), you would instead calculate $P(X \leq k - 0.5)$ (Normal approx)

$$P(X \geq 17) = P(X \geq 16.5)$$

Example: Normal approximation or Binomial distribution (1/2)

Example: Vaccinated people testing positive for Covid-19 (revisited)

About 25% of people that test positive for Covid-19 are vaccinated for it. Suppose 100 people have tested positive for Covid-19 (independently of each other). Let X denote the number of people that are vaccinated among the 100 that tested positive. What is the probability that fewer than 20 of the people that tested positive are vaccinated?

1. Calculate exact probability.
2. Calculate approximate probability.

$p = 0.25, n = 100$, we want $P(X < 20)$

X is # ppl vaccinated

1. Exact probability = Binomial distribution

$$X \sim \text{Binomial}(n = 100, p = 0.25)$$

$$P(X < 20) = P(X \leq 19) = \sum_{j=0}^{19} P(X = j)$$

$$P(X \leq 19) =$$

```
1 pbinom(q = 19, size = 100, prob = 0.25)
[1] 0.09953041
```

n p

Example: Normal approximation or Binomial distribution (2/2)

Example: Vaccinated people testing positive for Covid-19 (revisited)

About 25% of people that test positive for Covid-19 are vaccinated for it. Suppose 100 people have tested positive for Covid-19 (independently of each other). Let X denote the number of people that are vaccinated among the 100 that tested positive. What is the probability that fewer than 20 of the people that tested positive are vaccinated?

1. Calculate exact probability.
2. Calculate approximate probability.

need to show that
 $np > 10$ ✓ $n(1-p) > 10$ ✓
 $100 \cdot 0.25 = 25 > 10$ ✓ $100(1-0.25) = 75 > 10$ ✓
we can use Normal approx

$p = 0.25, n = 100$, we want $P(X < 20)$

2. Approximate probability = Normal distribution

- Mean = $\mu = np = 0.25 \cdot 100 = 25$
- SD = $\sigma = \sqrt{np(1-p)} = \sqrt{100 \cdot 0.25 \cdot (1-0.25)} = 4.33$

$$X \sim \text{Normal}(\mu = 25, \sigma = 4.33)$$

- Use **continuity correction**: Instead of calculating $P(X \leq 19)$, we calculate $P(X \leq 19.5)$

```
1 pnorm(q = 19.5, mean = 25,  
2      sd = sqrt(100*0.25*0.75))
```

```
[1] 0.1020119
```

$$0.102 \approx 0.0995$$

Some resources for the normal distribution

- [Page on R commands](#)