American University of Beirut
STAT 210: Elementary Statistics for Sciences
2022–2023 Fall

Siamak Taati

Chapter 6
Continuous random variables
Part 1

Continuous random variables

Random variables

A random variable within a random experiment is a variable X (often numerical) whose value is determined by the outcome of the experiment.

Two types of random variables

- A discrete random variable is one which has only finitely many or countably many possible values.

  For example:

  $\longrightarrow$ Number of heads in two flips a coin
  $\longrightarrow$ Sum of the numbers on two dice
  $\longrightarrow$ Number of spam emails during November

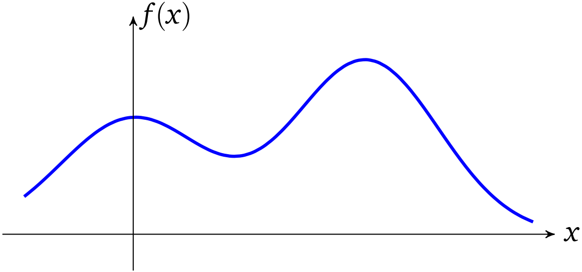
- A continuous random variable is one whose values range over an entire interval (or intervals).

  For example:

  $\longrightarrow$ Time until the next spam email
  $\longrightarrow$ Tomorrow's temperature in Beirut

Continuous random variables

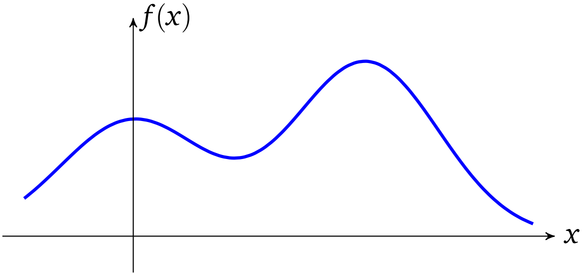
The distribution of a continuous RV X can be described by an idealized (density-type) histogram.



This can be thought of as the histogram of X in many many repeated experiments.

Continuous random variables

The distribution of a continuous RV X can be described by an idealized (density-type) histogram.

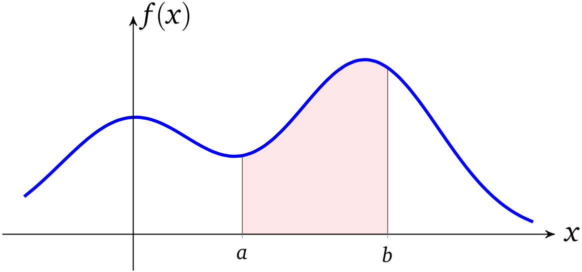


This can be thought of as the histogram of X in many many repeated experiments.

The vertical axis is normalized in such a way that the total area under the histogram is 1.

Continuous random variables

The distribution of a continuous RV X can be described by an idealized (density-type) histogram.

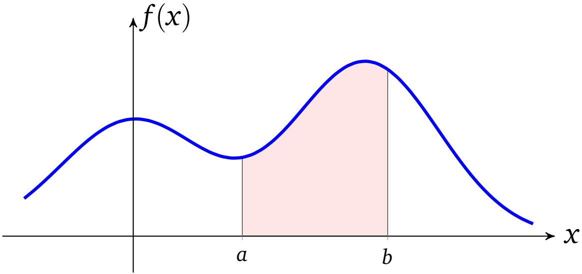


For every two points a and b,

$$\mathbb{P}(a \leq X \leq b) = \text{area of the shaded region}$$

Continuous random variables

The distribution of a continuous RV X can be described by an idealized (density-type) histogram.

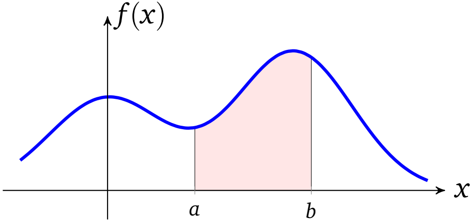


For every two points a and b,

$$\mathbb{P}(a \leq X \leq b) = \text{area of the shaded region}$$

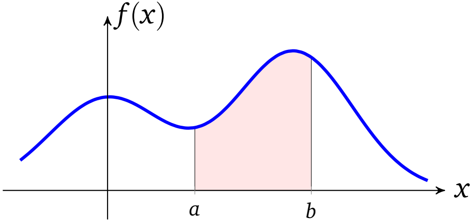
The blue curve is called the probability density function of X and is often denoted by $f(x)$.

Continuous random variables



$\mathbb{P}(a \leq X \leq b) =$ area of the shaded region

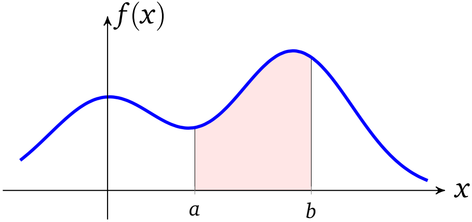
Continuous random variables



$\mathbb{P}(a \leq X \leq b) =$ area of the shaded region

Q What is $\mathbb{P}(X = a)$?

Continuous random variables



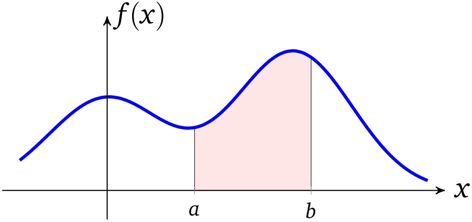
$\mathbb{P}(a \le X \le b) =$ area of the shaded region

Q What is $\mathbb{P}(X = a)$?

A Imagine $b = a$.

$$\mathbb{P}(X = a) = \mathbb{P}(a \le X \le a) = \text{area of the shaded region} = \boxed{0}\ .$$

Continuous random variables



$\mathbb{P}(a \leq X \leq b) =$ area of the shaded region

Q What is $\mathbb{P}(X = a)$?

A Imagine $b = a$.

$$\mathbb{P}(X = a) = \mathbb{P}(a \leq X \leq a) = \text{area of the shaded region} = \boxed{0}\,.$$

In general:

The probability that a <u>continuous RV</u> takes any given value is 0.

Continuous random variables

Some important types of continuous RVs:

- Normal RVs
- Uniform RVs (not covered in this course)
- Exponential RVs (not covered in this course)
- Gamma RVs (not covered in this course)
- t RVs
- chi-squared RVs
- F RVs
- ...

These are identified by their distributions.

Continuous random variables

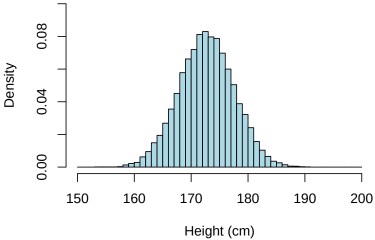
Some important types of continuous RVs:

- Normal RVs
- Uniform RVs (not covered in this course)
- Exponential RVs (not covered in this course)
- Gamma RVs (not covered in this course)
- t RVs
- chi-squared RVs
- F RVs
- . . .

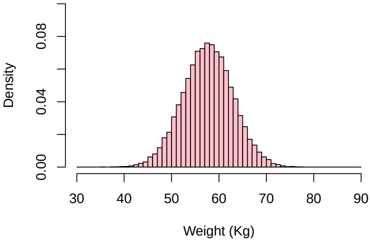
These are identified by their distributions.

Normal RVs

Recall the bell-shaped histograms of the heights and weights of 25,000 18-year-old individuals from Hong-Kong.



Histogram of the height



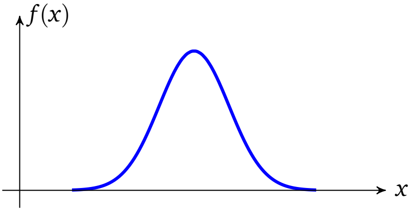
Histogram of the weight

The normal distribution is an idealization of such histograms.

It shows up in Statistics in various contexts, both as a model of a wide range of random quantities, and as the (derived) distribution of certain sample statistics.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

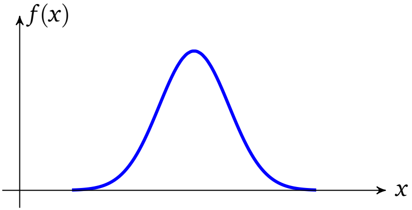


$$f(x) := \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A continuous RV with a normal distribution is called a normal RV.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:



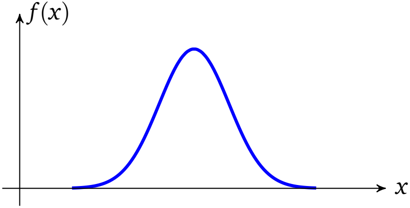
$$f(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A continuous RV with a normal distribution is called a normal RV.

The normal distribution is symmetric and unimodal. Its right and left tails extend indefinitely, which is to say:

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:



$$f(x) := \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

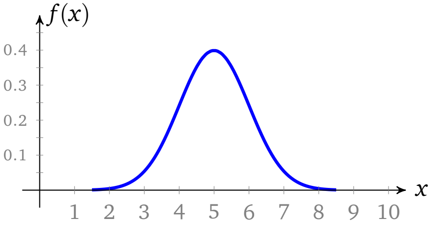
A continuous RV with a normal distribution is called a normal RV.

The normal distribution is <u>symmetric</u> and <u>unimodal</u>. Its right and left tails extend indefinitely, which is to say:

> The possible values of a normal RV are <u>all numbers</u>
> $-\infty < x < +\infty$.

Normal RVs

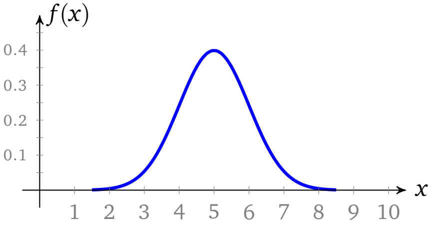
The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:



The meaning of the parameters:

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

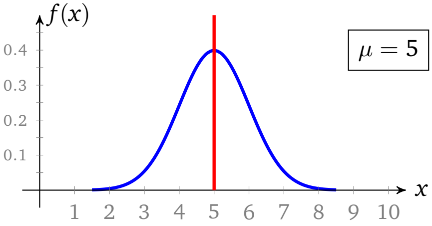


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

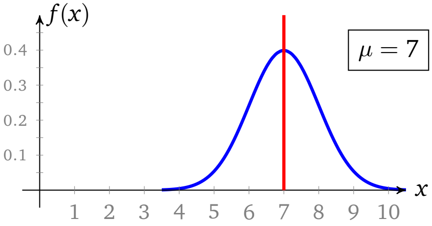


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

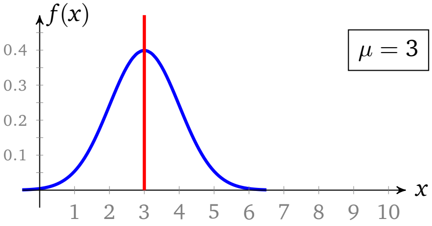


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

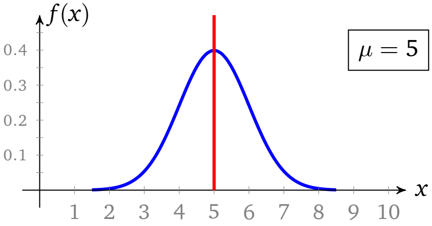


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

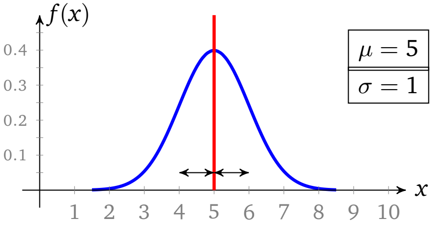


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.
- σ is the standard deviation. It controls the spread of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

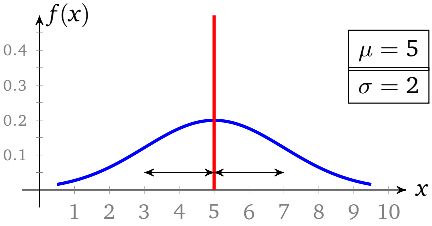


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.
- σ is the standard deviation. It controls the spread of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

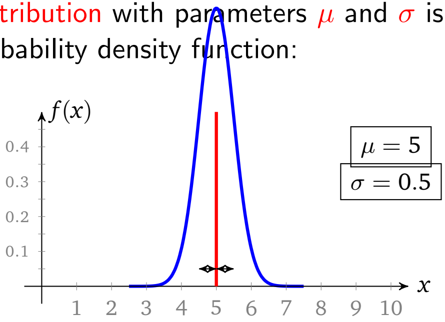


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.
- σ is the standard deviation. It controls the spread of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:

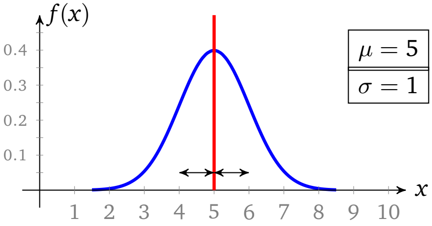


The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.
- σ is the standard deviation. It controls the spread of the distribution.

Normal RVs

The normal distribution with parameters μ and σ is a specific bell-shaped probability density function:



The meaning of the parameters:

- μ is the mean. It controls the location of the center of the distribution.
- σ is the standard deviation. It controls the spread of the distribution.

Normal RVs

Example (Height of a random 18-year old)

Estimates obtained from a statistical sample suggest that among the 18-year-old individuals from Lebanon, the population mean is 173.2 cm and the population standard deviation is 4.85 cm.

Normal RVs

Example (Height of a random 18-year old)

Estimates obtained from a statistical sample suggest that among the 18-year-old individuals from Lebanon, the population mean is 173.2 cm and the population standard deviation is 4.85 cm.

Let H denote the height of a randomly picked 18-year-old individual from Lebanon. Analogy with the data from Hong-Kong suggests that H is approximately normal.

Normal RVs

Example (Height of a random 18-year old)

Estimates obtained from a statistical sample suggest that among the 18-year-old individuals from Lebanon, the population mean is 173.2 cm and the population standard deviation is 4.85 cm.

Let H denote the height of a randomly picked 18-year-old individual from Lebanon. Analogy with the data from Hong-Kong suggests that H is <u>approximately</u> normal.

Q What are the parameters of this normal distribution?

Normal RVs

Example (Height of a random 18-year old)

Estimates obtained from a statistical sample suggest that among the 18-year-old individuals from Lebanon, the population mean is 173.2 cm and the population standard deviation is 4.85 cm.

Let H denote the height of a randomly picked 18-year-old individual from Lebanon. Analogy with the data from Hong-Kong suggests that H is <u>approximately</u> normal.

Q What are the parameters of this normal distribution?

$\boxed{\text{A}}$ $\mu = 173.2\,\text{cm}$ and $\sigma = 4.85\,\text{cm}$.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q1 What is the chance that the picked individual is shorter than 180 cm?

Normal RVs

Example (Height of a random 18-year old)

$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q1 What is the chance that the picked individual is shorter than 180 cm?

$\boxed{\text{A}}$ $\mathbb{P}(H < 180)$ is the area of the region under the probability density function to the left of 180.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q1 What is the chance that the picked individual is shorter than 180 cm?

$\boxed{\text{A}}$ $\mathbb{P}(H < 180)$ is the area of the region under the probability density function to the left of 180.

Using the applet for normal distribution

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q1 What is the chance that the picked individual is shorter than 180 cm?

A $\mathbb{P}(H < 180)$ is the area of the region under the probability density function to the left of 180.

Using the applet for normal distribution

$$\mathbb{P}(H < 180) = \text{area of the region to the left of } 180$$
$$\approx \boxed{0.91955} \approx 92.0\%$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q2 What is the chance that the picked individual is taller than $165\,\text{cm}$?

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q2 What is the chance that the picked individual is taller than 165 cm?

$\boxed{\text{A}}$ Using the applet for normal distribution

$$\mathbb{P}(H > 165) = \text{area of the region to the right of 160}$$
$$\approx \boxed{0.95456} \approx 95.5\%$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q3 What is the chance that the height of the picked individual is between $165\,\text{cm}$ and $180\,\text{cm}$?

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q3 What is the chance that the height of the picked individual is between $165\,\text{cm}$ and $180\,\text{cm}$?

$\boxed{\text{A3.1}}$ Using the applet for normal distribution

$$\begin{aligned}
\mathbb{P}(165 < H < 180) &= \text{area of the region between 165 and 180} \\
&= (\text{area to the left of 180}) - \\
&\qquad (\text{area to the left of 165}) \\
&\approx 0.91955 - 0.04544 \\
&= \boxed{0.87411} \approx 87.4\%
\end{aligned}$$

Normal RVs

Example (Height of a random 18-year old)

$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q3 What is the chance that the height of the picked individual is between $165\,\text{cm}$ and $180\,\text{cm}$?

A3.2 Using what we already have:

$$\begin{aligned}
\mathbb{P}(165 < H < 180) &= \mathbb{P}(H < 180) + \mathbb{P}(H > 165) - 1 \\
&\approx 0.91955 + 0.95456 - 1 \\
&= \boxed{0.87411} \approx 87.4\%
\end{aligned}$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q4 What is the chance that the picked individual has height $173.2\,\text{cm}$?

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q4 What is the chance that the picked individual has height $173.2\,\text{cm}$?

A $\mathbb{P}(H = 173.2) = \boxed{0}$.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q5 What is the chance that the picked individual has height at least $173.2\,\text{cm}$?

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q5 What is the chance that the picked individual has height at least $173.2\,\text{cm}$?

A By symmetry,

$$\mathbb{P}(H \geq \mu) = \mathbb{P}(H \leq \mu) = \boxed{1/2}$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q6 Find x such that the chance that the picked individual is shorter than x is 80%.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q6 Find x such that the chance that the picked individual is shorter than x is 80%.

A We want x such that

$$0.8 = \mathbb{P}(H < x) = \text{area of the region to the left of } x$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q6 Find x such that the chance that the picked individual is shorter than x is 80%.

A We want x such that

$$0.8 = \mathbb{P}(H < x) = \text{area of the region to the left of } x$$

Using the applet for normal distribution, $x \approx \boxed{177.28\,\text{cm}}$.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q7 Find x such that with probability 90%, the height of the picked individual is between $\mu - x$ and $\mu + x$.

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle\text{height of a randomly picked 18-year-old from Lebanon}\rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q7 Find x such that with probability 90%, the height of the picked individual is between $\mu - x$ and $\mu + x$.

$\boxed{\text{A}}$ We want x such that

$$\begin{aligned} 0.9 &= \text{area between } \mu - x \text{ and } \mu + x \\ &= 1 - \big(\text{area to the left of } \mu - x\big) - \big(\text{area to the right of } \mu + x\big) \end{aligned}$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q7 Find x such that with probability 90%, the height of the picked individual is between $\mu - x$ and $\mu + x$.

$\boxed{\text{A}}$ We want x such that

$$\begin{aligned} 0.9 &= \text{area between } \mu - x \text{ and } \mu + x \\ &= 1 - \big(\underbrace{\text{area to the left of } \mu - x}_{p}\big) - \big(\underbrace{\text{area to the right of } \mu + x}_{p}\big) \end{aligned}$$

By <u>symmetry</u>, the two areas are the same, hence

$$0.9 = 1 - 2p \qquad \Longrightarrow \qquad p = 0.05\ .$$

Normal RVs

Example (Height of a random 18-year old)

$$\boxed{H := \langle \text{height of a randomly picked 18-year-old from Lebanon} \rangle}$$

H is a normal RV with mean $\mu = 173.2\,\text{cm}$ and standard deviation $\sigma = 4.85\,\text{cm}$.

Q7 Find x such that with probability 90%, the height of the picked individual is between $\mu - x$ and $\mu + x$.

A We want x such that

$$\begin{aligned} 0.9 &= \text{area between } \mu - x \text{ and } \mu + x \\ &= 1 - \underbrace{\left(\text{area to the left of } \mu - x\right)}_{p} - \underbrace{\left(\text{area to the right of } \mu + x\right)}_{p} \end{aligned}$$

By <u>symmetry</u>, the two areas are the same, hence

$$0.9 = 1 - 2p \qquad \Longrightarrow \qquad p = 0.05\ .$$

Using the applet for normal distribution, $\mu - x \approx 165.22246$, which gives $x \approx \boxed{7.97754\,\text{cm}}$.