Week 8: Random variables & Sampling

Elena & Willa
10/19/2020

Today’s agenda

- Warm-up
- Random variables and distributions
- Sampling distributions

Random variables

So far, we have focused on directly assigning values to variables.

```
num <- 1

df <- penguins.csv

names <- c(“Willa”, “Elena”, “Roya”)
```

Random variables

So far, we have focused on directly assigning values to variables.

```
num <- 1

df <- penguins.csv

names <- c("Willa", "Elena", "Roya")
```

What is a random variable?

Berkeley
UNIVERSITY OF CALIFORNIA

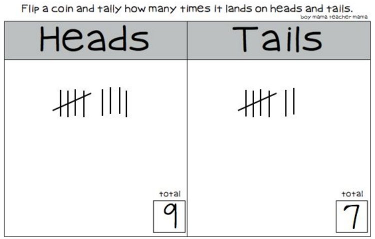
Random variables

“ A variable that takes on different numerical values as a result of a random experiment (eg. flipping a coin) or random measurements (eg. randomly sampling height in the population).”

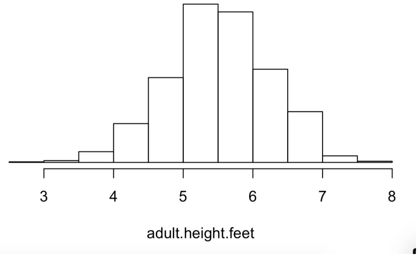
Random variables

" A variable that takes on different numerical values as a result of a random experiment (eg. flipping a coin) or random measurements (eg. randomly sampling height in the population). "

Randomly tossing a coin



Randomly sampling height in a population

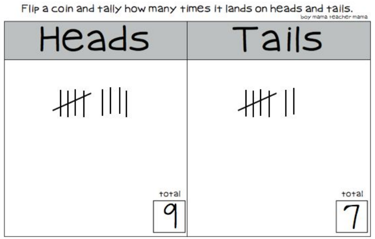


A random variable has a set of possible values, but the outcome each time is unknown.

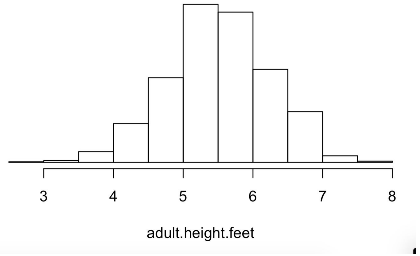
Random variables

" A variable that takes on different numerical values as a result of a random experiment (eg. flipping a coin) or random measurements (eg. randomly sampling height in the population).

Randomly tossing a coin



Randomly sampling height in a population

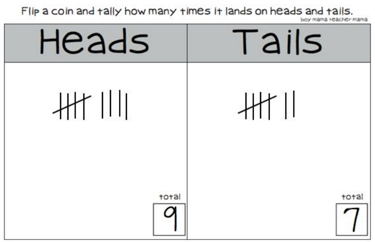


A random variable has a set of possible values, but the outcome each time is unknown.
As scientists, we spend our careers collecting measurements from random variables.

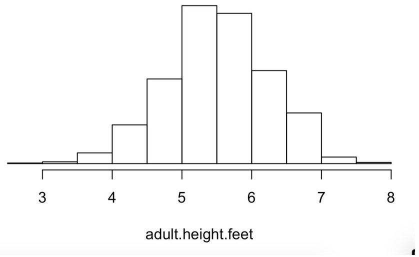
Random variables

" A variable that takes on different numerical values as a result of a random experiment (eg. flipping a coin) or random measurements (eg. randomly sampling height in the population).

Randomly tossing a coin



Randomly sampling height in a population



A random variable has a set of possible values, but the outcome each time is unknown.
As scientists, we spend our careers collecting measurements from random variables.

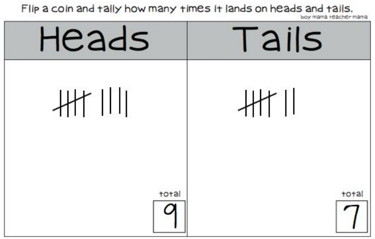
- Characterizing the distribution of the variable
- Statistical inference
- Predicting other outcomes from their values
- Comparing distributions amongst populations

Berkeley
UNIVERSITY OF CALIFORNIA

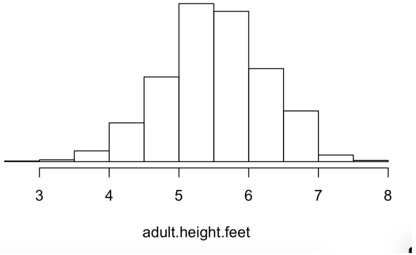
Random variables

" A variable that takes on different numerical values as a result of a random experiment (eg. flipping a coin) or random measurements (eg. randomly sampling height in the population).

Randomly tossing a coin



Randomly sampling height in a population



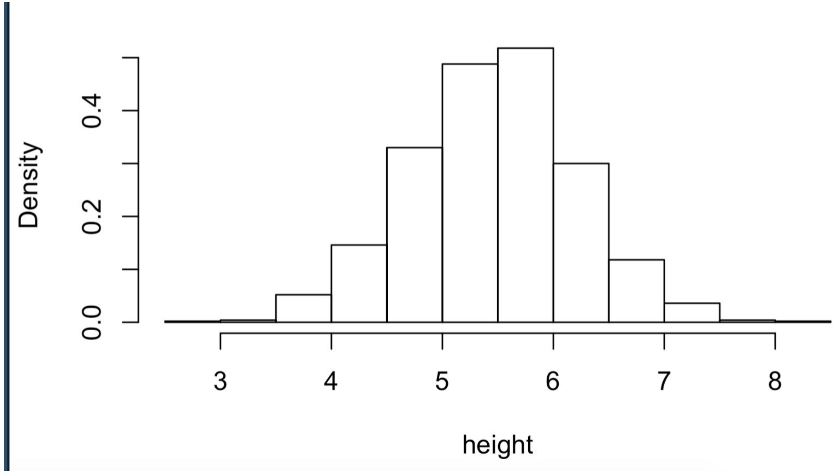
A random variable has a set of possible values, but the outcome each time is unknown.

As scientists, we spend our careers collecting measurements from random variables.

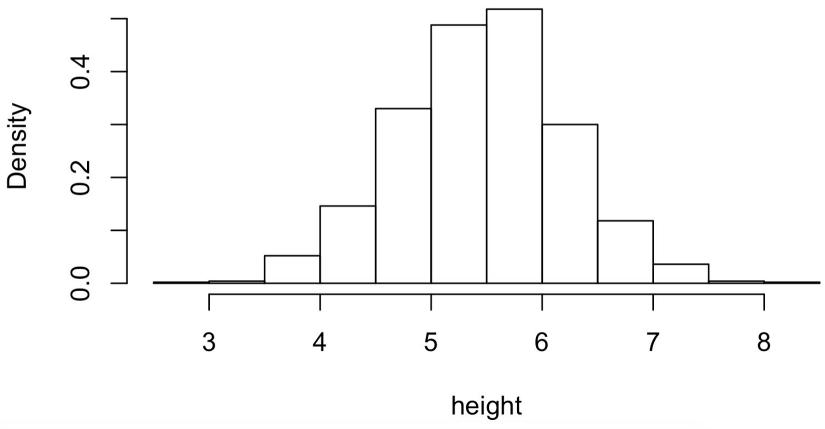
- Characterizing the distribution of the variable
- Statistical inference
- Predicting other outcomes from their values
- Comparing distributions amongst populations

Understanding the definition & properties of random variables becomes important when you are doing data analysis.

How do we define a random variable ?

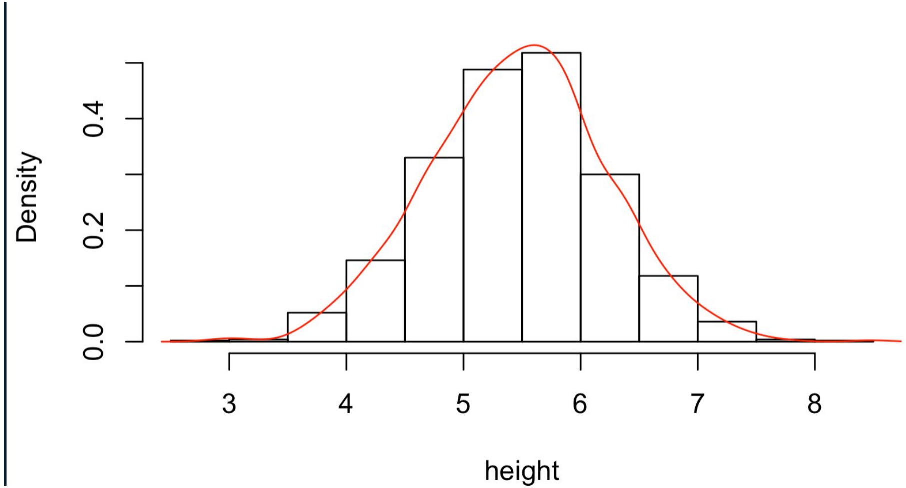


How do we define a random variable ?



- Set of possible values
- Each value has a probability of occurring.

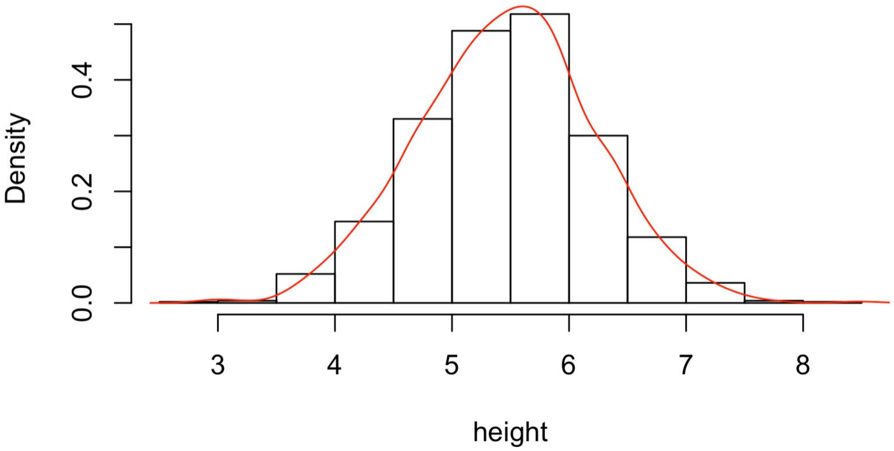
How do we define a random variable ?



Define our variable with a function (eg. normally distributed).

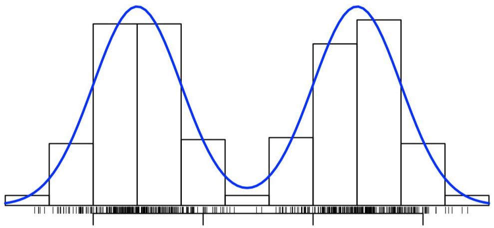
- Set of possible values
- Each value has a probability of occurring.

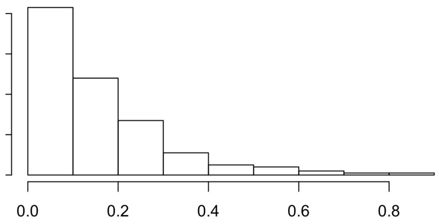
How do we define a random variable ?



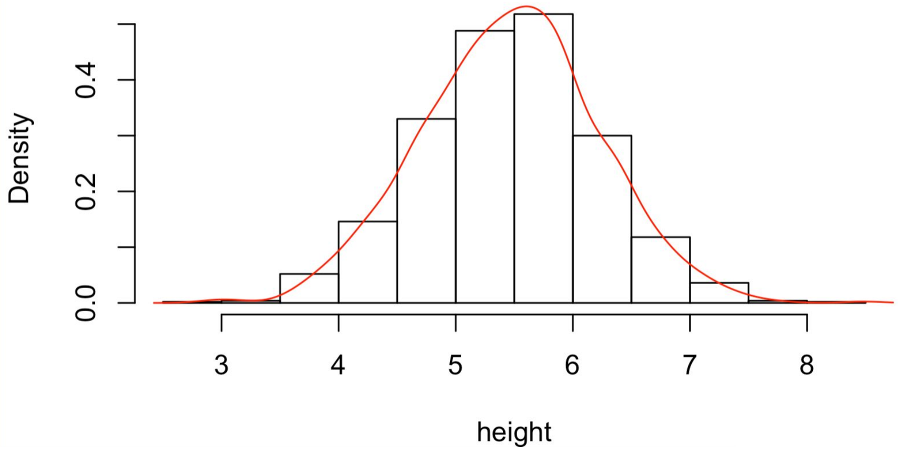
Define our variable with a function (eg. normally distributed).

- Set of possible values
- Each value has a probability of occurring.





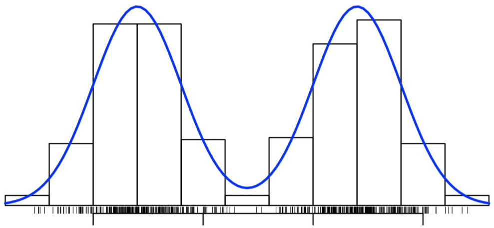
How do we define a random variable ?

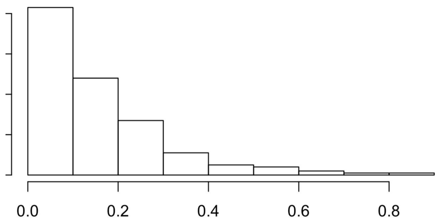


Define our variable with a function (eg. normally distributed).

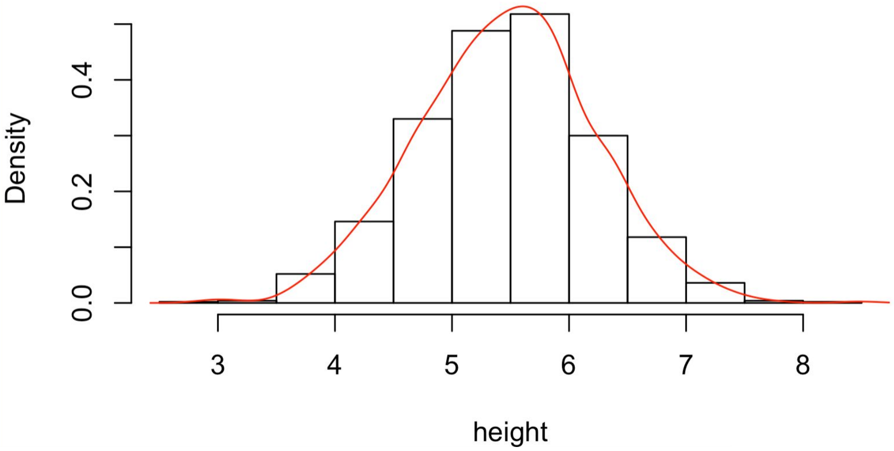
- Set of possible values
- Each value has a probability of occurring.

As scientists we often collect measurements without knowing the underlying distribution.



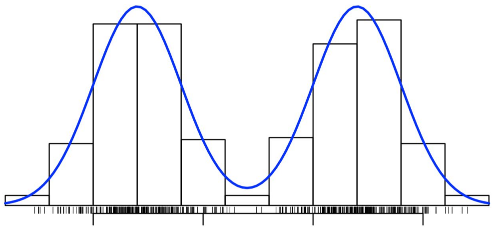


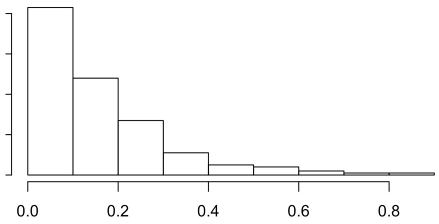
How do we define a random variable ?



Define our variable with a function (eg. normally distributed).

- Set of possible values
- Each value has a probability of occurring.

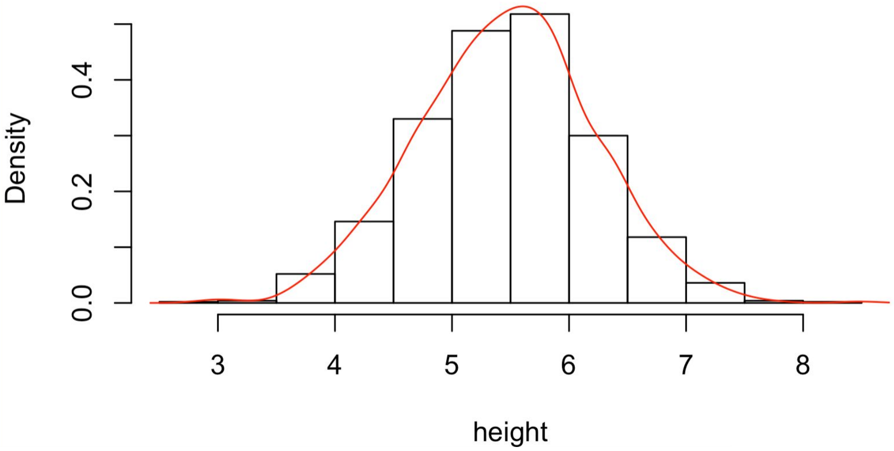




As scientists we often collect measurements without knowing the underlying distribution.

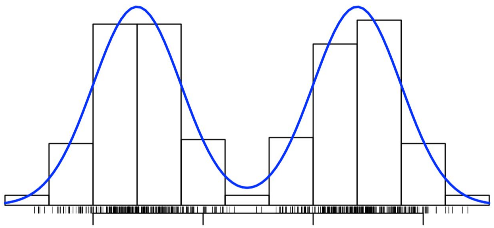
But we know each observation is one value of a range of possible outcomes.

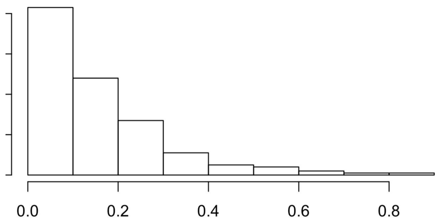
How do we define a random variable ?



Define our variable with a function (eg. normally distributed).

- Set of possible values
- Each value has a probability of occurring.



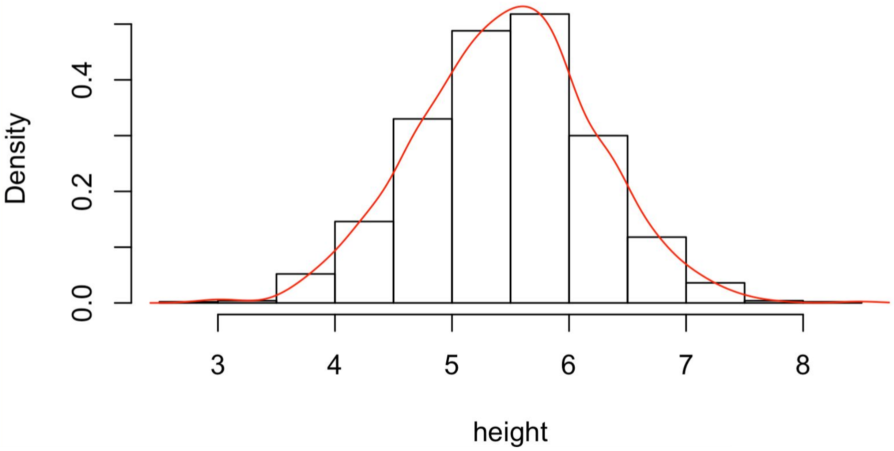


As scientists we often collect measurements without knowing the underlying distribution.

But we know each observation is one value of a range of possible outcomes.

We can represent the variable as a collection of our measured outcomes and then make inferences about the distribution.

How do we define a random variable ?



Define our variable with a function (eg. normally distributed).

- Set of possible values
- Each value has a probability of occurring.

As scientists we often collect measurements without knowing the underlying distribution.

But we know each observation is one value of a range of possible outcomes.

We can represent the variable as a collection of our measured outcomes and then make inferences about the distribution.

Thinking about your data this way can help make a lot of statistical techniques more intuitive!

Berkeley
UNIVERSITY OF CALIFORNIA

Simulating observations from a random variable

In R, we can simulate this data collection process and get observations for a random variable from a distribution.

Berkeley
UNIVERSITY OF CALIFORNIA

Simulating observations from a random variable

In R, we can simulate this data collection process and get observations for a random variable from a distribution.

In the real world we might not know the distribution, but in R we will define our random variable with a range of values and a probability function that we believe makes sense.

Simulating observations from a random variable

In R, we can simulate this data collection process and get observations for a random variable from a distribution.

In the real world we might not know the distribution, but in R we will define our random variable with a range of values and a probability function that we believe makes sense.

Then we will look at the range of values we can get from this variable.