Flipping a Penny

Exploring Randomness and Reviewing Nomenclature

David Housman

Goshen College

MathFest 2023

Randomness

- *Randomness* is "the quality or state of lacking a pattern or principle of organization" (Oxford Languages Online Dictionary).

Randomness

- *Randomness* is "the quality or state of lacking a pattern or principle of organization" (Oxford Languages Online Dictionary).
- Humans have evolved to discern patterns in the world.

Randomness

- *Randomness* is "the quality or state of lacking a pattern or principle of organization" (Oxford Languages Online Dictionary).
- Humans have evolved to discern patterns in the world.
- So, human intuition for randomness is not very good without training.

Randomness

- *Randomness* is "the quality or state of lacking a pattern or principle of organization" (Oxford Languages Online Dictionary).
- Humans have evolved to discern patterns in the world.
- So, human intuition for randomness is not very good without training.
- Flipping coins, rolling dice, and shuffling cards are usually thought of as random.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.
- We will ask for the number of heads and the length of the longest run of either side.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.
- We will ask for the number of heads and the length of the longest run of either side.
- *Data* are the answers given to the variable question by the members of a sample.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.
- We will ask for the number of heads and the length of the longest run of either side.
- *Data* are the answers given to the variable question by the members of a sample.
- The *distribution* displays the absolute or relative count of each possible answer from the data.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.
- We will ask for the number of heads and the length of the longest run of either side.
- *Data* are the answers given to the variable question by the members of a sample.
- The *distribution* displays the absolute or relative count of each possible answer from the data.
- Some ways to display a distribution include a frequency table, histogram, density plot, or box and whiskers plot.

Nomenclature

- A *population* is the set of all people, objects, or cases of interest.
- Our population is all ten air flips of a penny.
- A *sample* is a subset of the population.
- We will generate samples of this population via human simulation, actual penny flips, and computer simulation.
- A *variable* is a question that can be asked of members of the population.
- We will ask for the number of heads and the length of the longest run of either side.
- *Data* are the answers given to the variable question by the members of a sample.
- The *distribution* displays the absolute or relative count of each possible answer from the data.
- Some ways to display a distribution include a frequency table, histogram, density plot, or box and whiskers plot.
- Now how I might begin the class ....

Human Simulation

- Simulate air flipping a penny ten times.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.
- Repeat three or four more times.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.
- Repeat three or four more times.
- A *population* is the set of cases of interest.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible human simulated ten air flips of a penny.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible human simulated ten air flips of a penny.
- A *sample* is a subset of the population.

Human Simulation

- Simulate air flipping a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1000110101 is what I recorded for my simulation.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible human simulated ten air flips of a penny.
- A *sample* is a subset of the population.
- In our example, the sample is all of the human simulated ten air flips of a penny done by this class.

Actual Flips

- Air flip a penny ten times.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.
- Repeat three or four more times.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.
- Repeat three or four more times.
- A *population* is the set of cases of interest.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible actual ten air flips of a penny.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible actual ten air flips of a penny.
- A *sample* is a subset of the population.

Actual Flips

- Air flip a penny ten times.
- Record the outcome as a string of ten 1's (heads) and 0's (tails).
- For example, 1011110010 is what I recorded for my actual flips.
- Repeat three or four more times.
- A *population* is the set of cases of interest.
- In our example, we are interested in all possible actual ten air flips of a penny.
- A *sample* is a subset of the population.
- In our example, the sample is all of the actual ten air flips of a penny done by this class.

Computer Simulation via R

- Simulate air flipping a penny ten times.

Computer Simulation via R

- Simulate air flipping a penny ten times.
- sample(c(0,1), size = 10, replace = TRUE)
  [1] 1 1 0 1 0 0 1 1 1 0

Computer Simulation via R

- Simulate air flipping a penny ten times.
- sample(c(0,1), size = 10, replace = TRUE)
  [1] 1 1 0 1 0 0 1 1 1 0
- Repeat four times.

Computer Simulation via R

- Simulate air flipping a penny ten times.
- sample(c(0,1), size = 10, replace = TRUE)
  [1] 1 1 0 1 0 0 1 1 1 0
- Repeat four times.
- for (i in seq(4)) {
  print(sample(c(0,1), size = 10, replace = TRUE))
  }
  [1] 0 0 0 1 1 0 1 1 1 0
  [1] 0 1 0 1 1 0 1 1 0 1
  [1] 0 0 1 1 0 0 1 1 1 1
  [1] 0 1 0 0 1 1 1 1 1 1

Variables

- A *variable* is a question that can be asked of members of the population.

Variables

- A *variable* is a question that can be asked of members of the population.
- *Data* are the answers given to the variable question by the members of a sample.

Variables

- A *variable* is a question that can be asked of members of the population.
- *Data* are the answers given to the variable question by the members of a sample.
- For each of your cases, find the number of heads (heads) and length of the longest run of either heads or tails (run_length).

Variables

- A *variable* is a question that can be asked of members of the population.
- *Data* are the answers given to the variable question by the members of a sample.
- For each of your cases, find the number of heads (heads) and length of the longest run of either heads or tails (run_length).
- For example,

case	heads	run_length
0 0 0 1 1 0 1 1 1 0	5	3
0 1 0 1 1 0 1 1 0 1	6	2
0 0 1 1 0 0 1 1 1 1	6	4
1 0 1 1 0 0 0 0 0 0	3	6

Variables

- A *variable* is a question that can be asked of members of the population.
- *Data* are the answers given to the variable question by the members of a sample.
- For each of your cases, find the number of heads (heads) and length of the longest run of either heads or tails (run_length).
- For example,

case	heads	run_length
0 0 0 1 1 0 1 1 1 0	5	3
0 1 0 1 1 0 1 1 0 1	6	2
0 0 1 1 0 0 1 1 1 1	6	4
1 0 1 1 0 0 0 0 0 0	3	6

- Record the values of your variables in the class spreadsheets "human.csv" and "penny.csv" with column headers "head" and "run_length".

Variables in R

- case = sample(c(0,1), size = 10, replace = TRUE); case
  [1] 0 1 1 1 0 0 0 0 0 1

Variables in R

- case = sample(c(0,1), size = 10, replace = TRUE); case
  [1] 0 1 1 1 0 0 0 0 0 1
- heads = sum(case); heads
  [1] 4

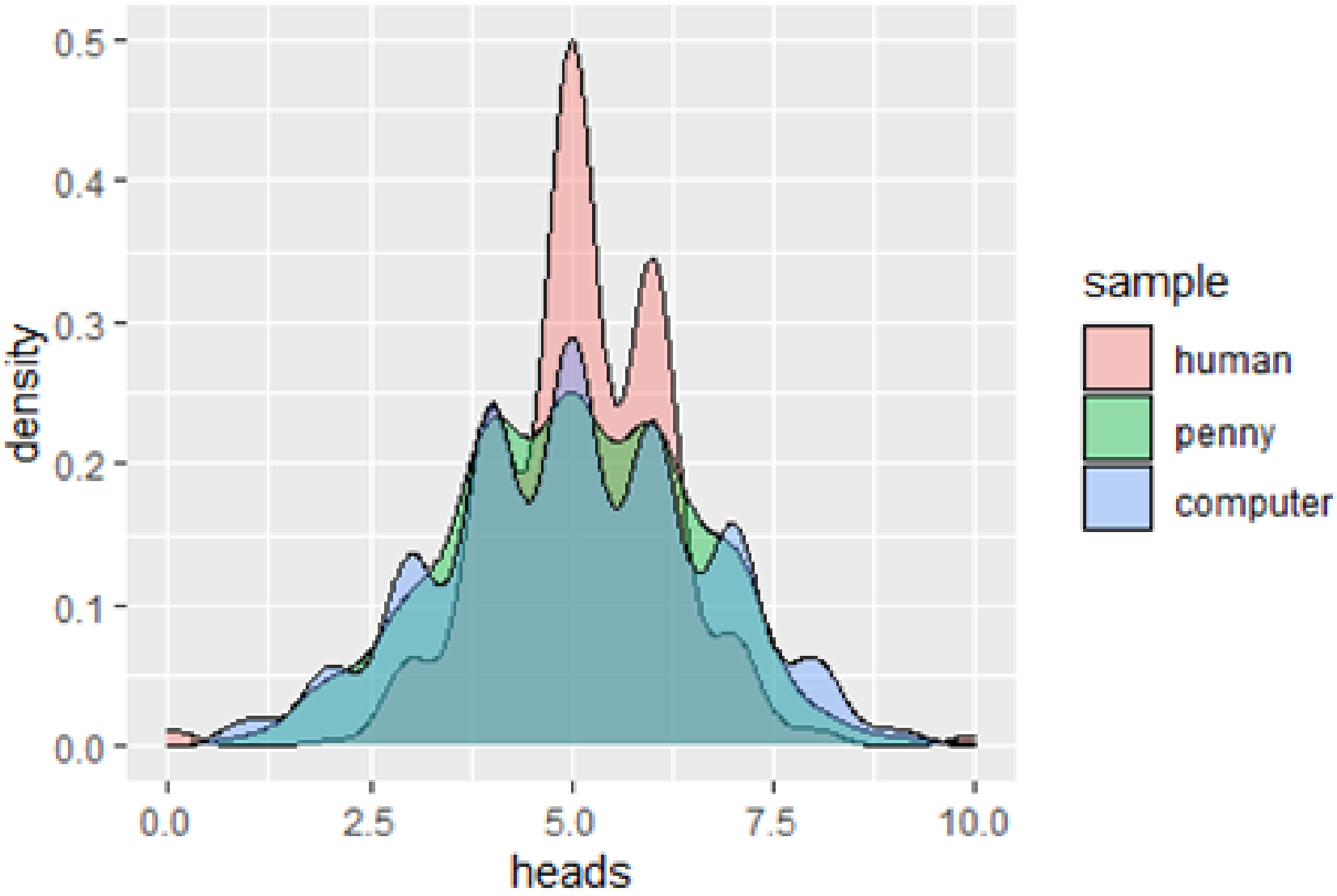
Variables in R

- case = sample(c(0,1), size = 10, replace = TRUE); case
  [1] 0 1 1 1 0 0 0 0 0 1
- heads = sum(case); heads
  [1] 4
- run_length = max(rle(case)$lengths); run_length
  [1] 5

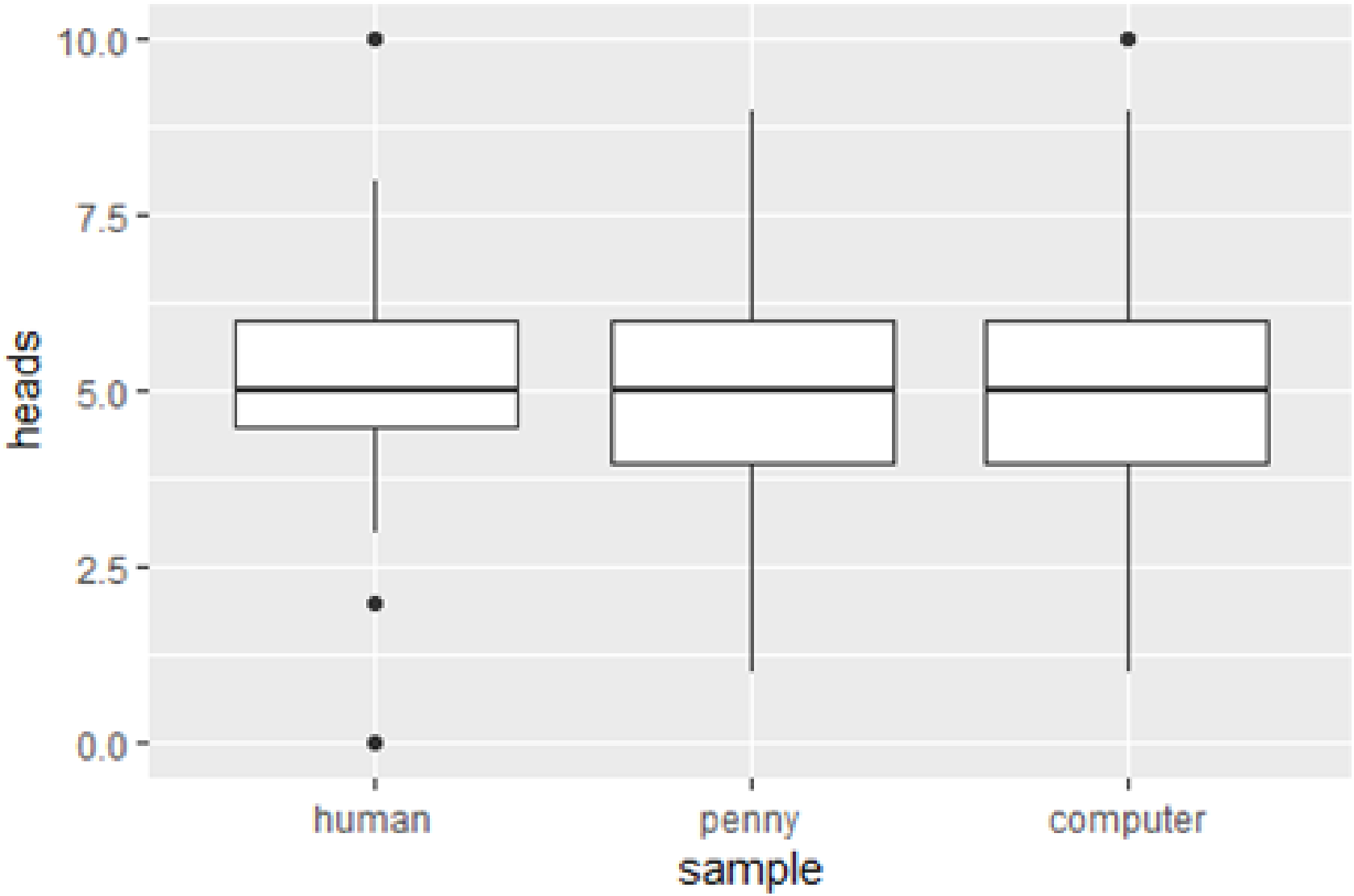
Variables in R

- case = sample(c(0,1), size = 10, replace = TRUE); case
  [1] 0 1 1 1 0 0 0 0 0 1
- heads = sum(case); heads
  [1] 4
- run_length = max(rle(case)$lengths); run_length
  [1] 5
- computer = tibble(heads = c(), run_length = c())
  for (i in seq(1000)) {
  case = sample(c(0,1), size = 10, replace = TRUE)
  one_row = tibble(
  heads = sum(case),
  run_length = max(rle(case)$lengths))
  computer = bind_rows(computer, one_row)
  }
  write_csv(computer, "computer.csv")

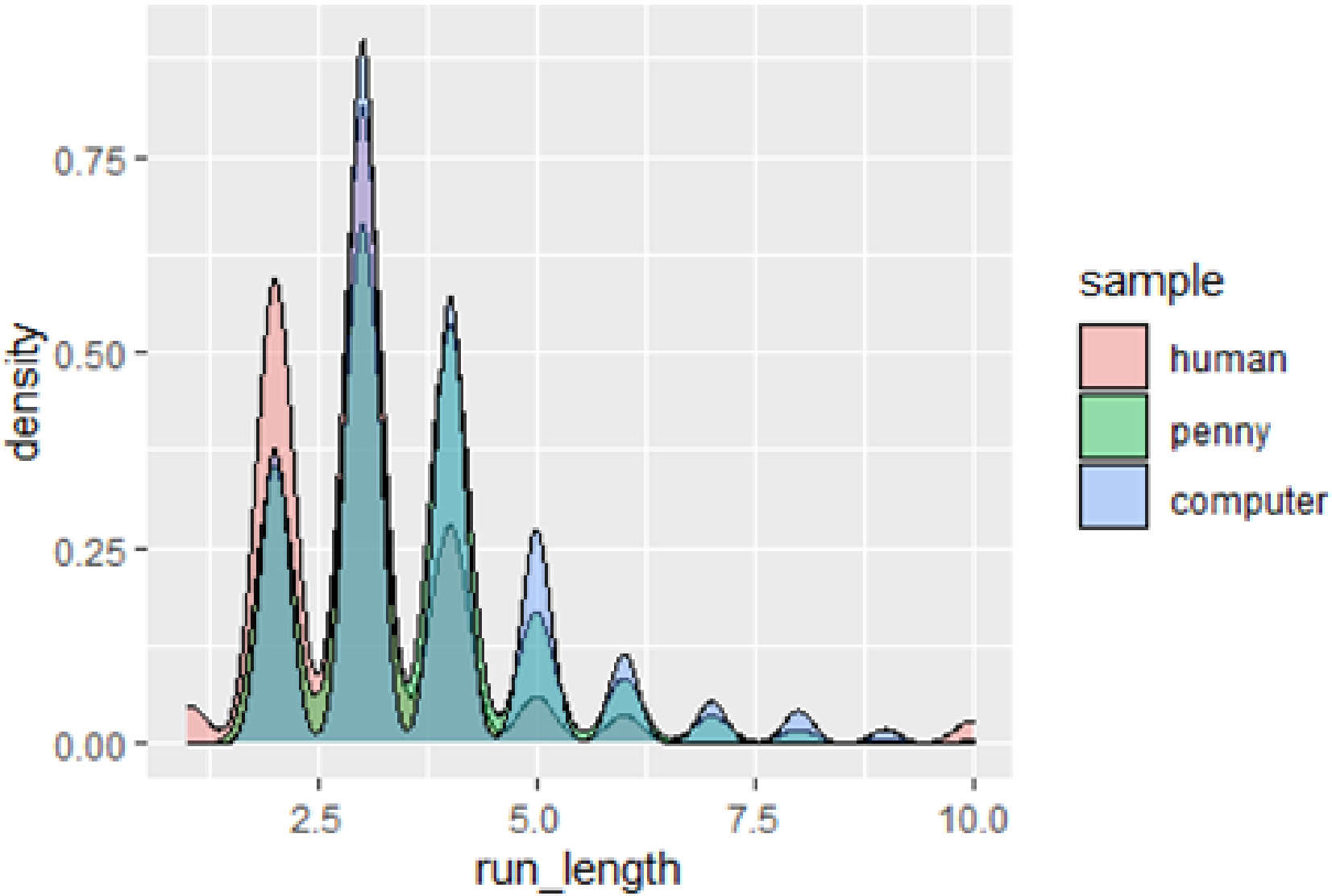
Number of Heads Comparison



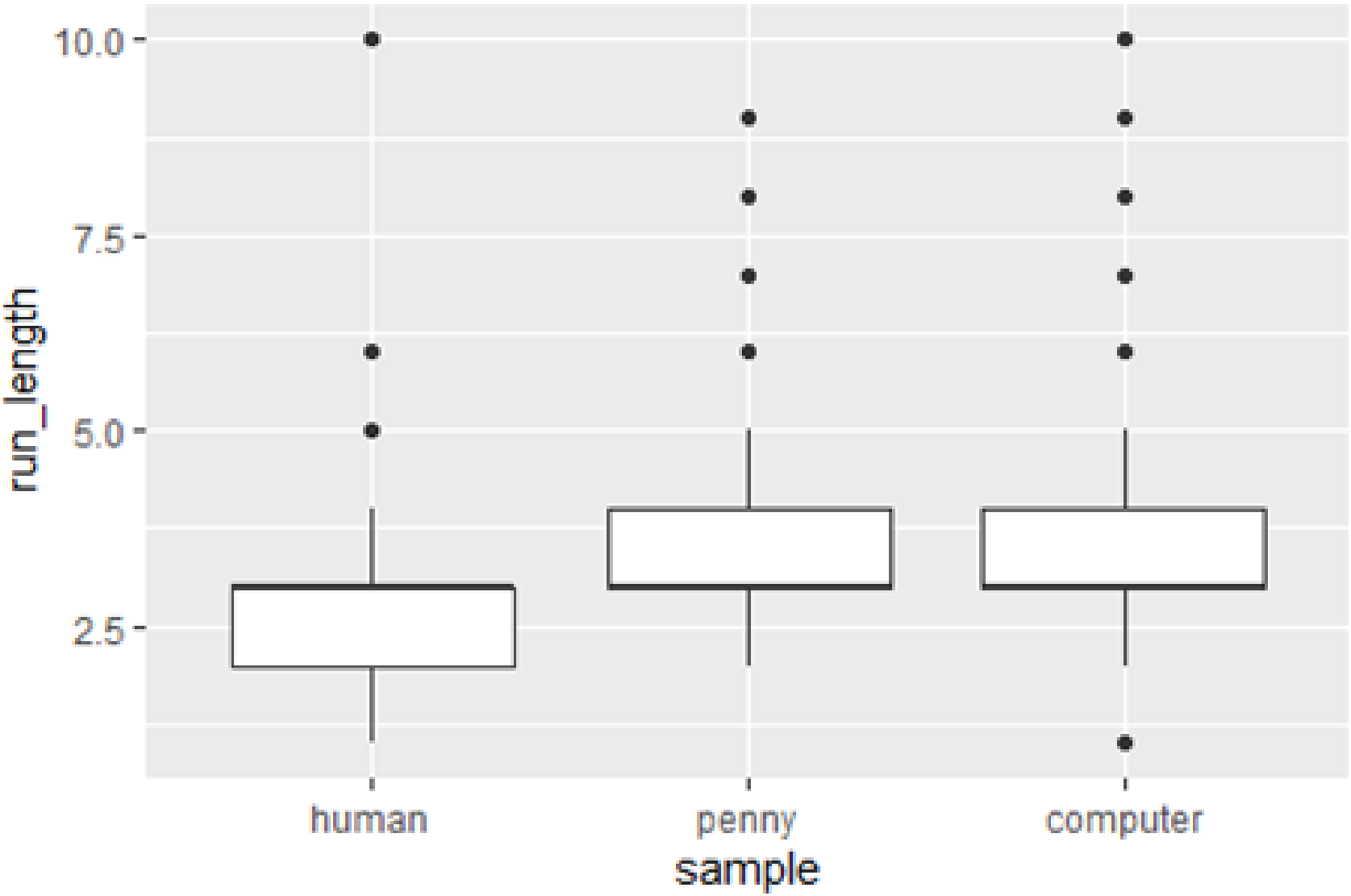
Number of Heads Comparison



Length of Longest Run Comparison



Length of Longest Run Comparison



Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.

Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.
- The activity introduces or reviews the concepts of population, sample, variable, data, distribution, and visualizations of distribution.

Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.
- The activity introduces or reviews the concepts of population, sample, variable, data, distribution, and visualizations of distribution.
- The higher level skill of graphical interpretation is practiced.

Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.
- The activity introduces or reviews the concepts of population, sample, variable, data, distribution, and visualizations of distribution.
- The higher level skill of graphical interpretation is practiced.
- The computer simulations can be done with any statistical software (even Excel) although finding run_length may be difficult.

Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.
- The activity introduces or reviews the concepts of population, sample, variable, data, distribution, and visualizations of distribution.
- The higher level skill of graphical interpretation is practiced.
- The computer simulations can be done with any statistical software (even Excel) although finding run_length may be difficult.
- Rather than the computer constructing comparison visualizations, students can construct the human simulation and actual data histograms by hand on a piece of graph paper that is projected for everyone to see.

Conclusions

- Simulating randomness accurately is difficult for humans but seemingly easy for deterministic computers.
- The activity introduces or reviews the concepts of population, sample, variable, data, distribution, and visualizations of distribution.
- The higher level skill of graphical interpretation is practiced.
- The computer simulations can be done with any statistical software (even Excel) although finding run_length may be difficult.
- Rather than the computer constructing comparison visualizations, students can construct the human simulation and actual data histograms by hand on a piece of graph paper that is projected for everyone to see.
- A fun followup activity is to have students "flip" a penny by first resting the penny on its side on a table and then jostling the table. After 100 or so flips have happened and the percentage of heads has been determined, ask students whether this is a fair way to flip a penny.