Dataframe basics

Bryan Clair

Wednesday, February 3, 2021

Statistical Experiments

Say something true about the world, even though we only observe a sample.

Key idea

- Introduce randomness into our observations and use **probability** to make precise statements.

A **statistical experiment** is a repeatable process that produces a random **outcome**.

The **sample space** is the set of all possible outcomes.

Sample Space

Sample space: the set of outcomes of a statistical experiment.

Flip a coin

$S = Heads, Tails$

Roll a die

$S = 1, 2, 3, 4, 5, 6$

Wait for a bus at a bus stop

$S = [0, \infty)$

Take a patient's blood pressure

S is pairs of positive numbers

Events

An event is a subset of the sample space.
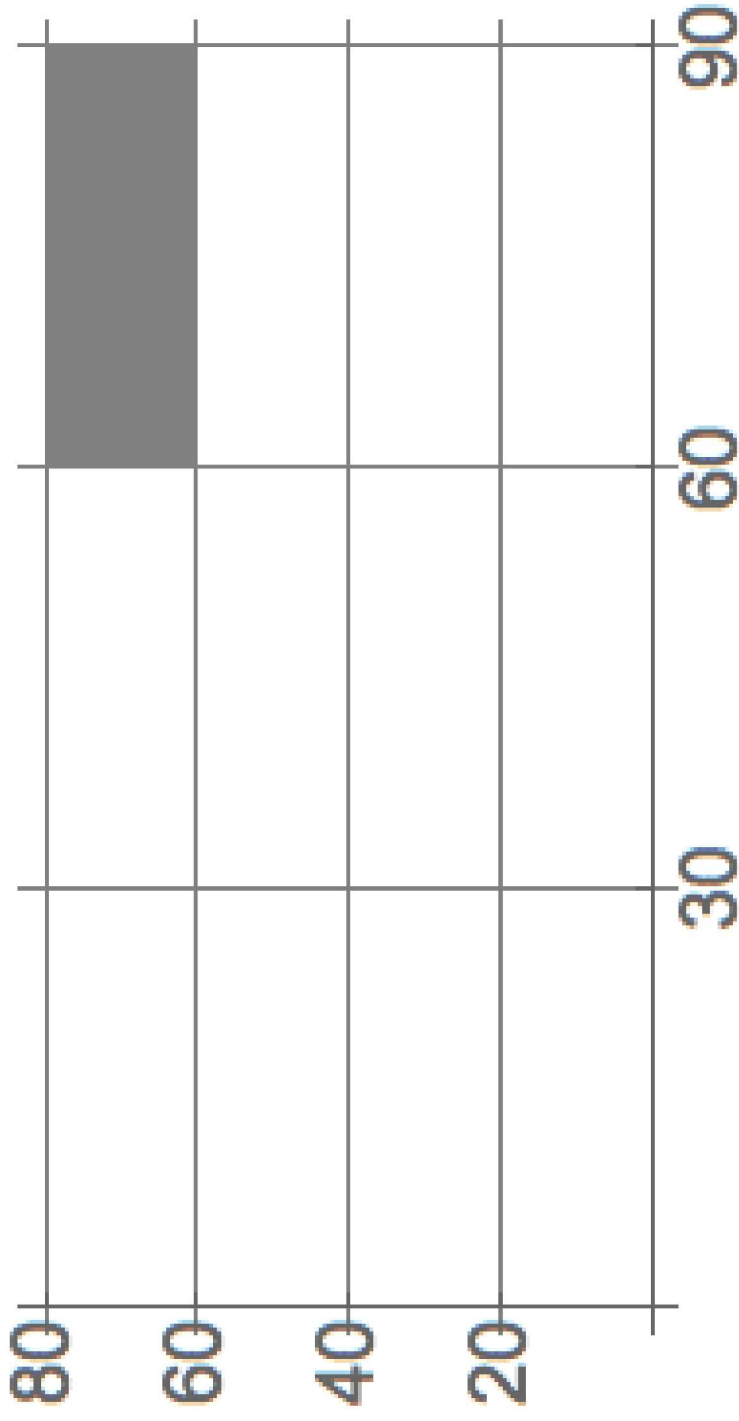
Flip a coin. There are four possible events:

$$\{\}, \{\text{Heads}\}, \{\text{Tails}\}, \{\text{Heads}, \text{Tails}\}$$

Roll a die. The event "even" is 2, 4, 6.

Wait for a bus. The event "wait less than five minutes" is $[0, 5)$.

Events

Check blood pressure. The event "normal blood pressure" is:

Probability

The probability of an event is a number between 0 and 1.

For an event A, write $P(A)$ for the probability.

$$0 \leq P(A) \leq 1$$

We *assign* probabilities to events in the hopes that they model the world.

The Law of Large Numbers:

$$\frac{\# \text{ of times } A \text{ occurs}}{\# \text{ of trials}} \rightarrow P(A)$$

as the number of trials of the experiment grows.

Simulation

We will use R to estimate probabilities by simulation.

`sample`: choose randomly from a vector

`replicate`: repeat an experiment

Examples

Roll a die. What is $P(\text{roll} < 3)$?

Roll two dice and sum them.

- What is $P(\text{roll} = 7)$?
- What is $P(\text{roll} = 12)$?
- What is $P(\text{roll is even})$?

Roll three dice and sum them. What is $P(\text{roll} > 12)$?

Flip 20 coins. What is the probability of getting exactly 10 heads?

Sample the lengths of four rivers from the `rivers` vector. What is the probability that your sample average is larger than 1000?

More Examples

10% of humans are left handed. Do we have an unusual number of lefties in this class?

If a 75% free throw shooter takes 10 free throws, what is the probability they make them all?

What is the probability they make more than half?