Probability prepare

STA 101

Getting started

Download this prepare file by pasting the code below into your **console** (bottom left of screen)

```
download.file("https://sta101-fa22.netlify.app/static/appex/prepareProbability.qmd",
  destfile = "prepareProbability.qmd")
```

Goals

- be able to define and give examples of sample space, outcomes, events, probabilities, conditional, marginal, joint and independent probabilities

Load packages

```
library(tidyverse)
library(fivethirtyeight)
```

Notes

Sample space

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

Discrete examples

- Experiment 1: You flip a coin once. The sample space is $\{H, T\}$.

We separate each outcome by a comma and use brackets {} to denote a "set".

- Experiment 2: You flip a coin twice. The sample space is $\{HH, HT, TH, TT\}$
- Experiment 3: You roll a die once. The sample space is $\{1, 2, 3, 4, 5, 6\}$
- Experiment 4: You send out a survey asking participants whether they prefer cats or dogs. The sample space is $\{\text{Cats}, \text{Dogs}\}$
- Experiment 5: A car manufacturer makes 100 vehicles. You count the number of recalls. The sample space is $\{0, 1, 2, 3, \ldots, 99, 100\}$

Continuous examples

- Experiment 6: You observe the numeric grade you earn in a course. The sample space is $[0, 100]$

Here we write the lower bound and upper bound of the sample space and assume we can observe all values in-between. Brackets, [ ], are inclusive of the end values while parentheses, ( ), are not.

- Experiment 7: You measure the tail length of American alligators The sample space is $(0, c]$ feet where c is the maximum tail length of an alligator, e.g. c might be approximately 10.
- Experiment 8: You measure the geographic coordinates (longitude and latitude) of a COVID case. The sample space is $[-90, 90]$ for latitude and $[-180, 180]$ for longitude.

Events

An **event** is a collection of 1 or more outcomes. Two events are said to be **disjoint** if they cannot occur at the same time.

Examples

- You roll a die once. Let A be the event that you roll an even number, i.e. $A = \{2, 4, 6\}$. Let B be the event you roll a 1 or a 2, i.e. $B = \{1, 2\}$. A and B are **not** disjoint.
- A car manufacturer makes 100 vehicles. You count the number of recalls. Let C be the event you see fewer than 10 recalls. $C = \{0, 1, 2, 3, ..., 8, 9\}$
- You observe the numeric grade you earn in a course. Let D be the event you receive a letter grade of "A". $D = [93, 100]$. Let E be the event that you earn a "B" or worse. $E = [0, 87)$. D and E **are disjoint** events because they cannot occur simultaneously.

Random variables

Random variables are functions that map outcomes to numbers. An **indicator random variable** takes values *1* and *0* to indicate whether or not an event occurs.

```
data(bob_ross) # within fivethirtyeight package
bob_ross %>%
  head(10)
```

```
# A tibble: 10 x 71
   episode season episode_num title      apple_frame aurora_borealis  barn beach
   <chr>    <dbl>       <dbl> <chr>            <int>           <int> <int> <int>
 1 S01E01       1           1 A WALK IN~           0               0     0     0
 2 S01E02       1           2 MT. MCKIN~           0               0     0     0
 3 S01E03       1           3 EBONY SUN~           0               0     0     0
 4 S01E04       1           4 WINTER MI~           0               0     0     0
 5 S01E05       1           5 QUIET STR~           0               0     0     0
 6 S01E06       1           6 WINTER MO~           0               0     0     0
 7 S01E07       1           7 AUTUMN MO~           0               0     0     0
 8 S01E08       1           8 PEACEFUL ~           0               0     0     0
 9 S01E09       1           9 SEASCAPE             0               0     0     1
10 S01E10       1          10 MOUNTAIN ~           0               0     0     0
# ... with 63 more variables: boat <int>, bridge <int>, building <int>,
#   bushes <int>, cabin <int>, cactus <int>, circle_frame <int>, cirrus <int>,
#   cliff <int>, clouds <int>, conifer <int>, cumulus <int>, deciduous <int>,
#   diane_andre <int>, dock <int>, double_oval_frame <int>, farm <int>,
#   fence <int>, fire <int>, florida_frame <int>, flowers <int>, fog <int>,
#   framed <int>, grass <int>, guest <int>, half_circle_frame <int>,
#   half_oval_frame <int>, hills <int>, lake <int>, lakes <int>, ...
```

One often writes indicator random variables as a bold "1",

$$\mathbf{1}_{\text{clouds}} = \begin{cases} 1 \text{ if there are clouds,} \\ 0 \text{ if not} \end{cases}$$

Probability

A **probability** is the long-run frequency of an *event*. In other words, the proportion of times we would see an event occur if we could repeat an experiment an infinite number of times. Probabilities take values between 0 and 1 inclusive.

- We can often compute probabilities practically as the mean of an indicator random variable. For example,

$$P(\text{clouds}) = \text{mean}(\mathbf{1}_{\text{clouds}})$$

- If A and B are two disjoint events, then the probability of A or B occurring is equal to the probability of A plus the probability of B. More concisely, $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$.

More definitions

Let A and B be two events.

- Marginal probability: The probability an event occurs regardless of values of the other event
  - $\text{P}(A)$
  - Example: What's the probability that, in a randomly selected episode of Bob Ross, the painting features clouds?
- Joint probability: The probability two or more events simultaneously occur
  - Example: What's the probability that, in a randomly selected episode of Bob Ross, the painting features clouds and mountains?
  - $\text{P}(A \text{ and } B)$
- Conditional probability: The probability an event occurs given the other has occurred
  - $\text{P}(A|B)$ or $\text{P}(B|A)$
  - Example: What is the probability that a Bob Ross painting features clouds in season 1?
  - $\text{P}(A|B) = \text{P}(A \text{ and } B)\ /\ \text{P}(B)$

- Independent events: Knowing one event has occurred does not lead to any change in the probability we assign to another event.
  - $\mathrm{P}(A|\ B) = \mathrm{P}(A)$ or $\mathrm{P}(B|A) = \mathrm{P}(B)$
  - Example: P(lakes | rivers) = P(lakes)