STA130H1S TUT0109
W4: Statistical Inference

Feb 1, 2019

Email: sonia.chhay@mail.utoronto.ca
Website: soniachhay.github.io

Announcements

- Be sure to answer all parts of the question, including the explanation if it asks for it
- You’re allowed to discuss assignment with other students but you cannot share your answers, anything you submit must be your own work

Overview

- Feedback
  - About the course
  - Course design
  - Communicating your work
  - Writing example (1) - Abstract
  - Writing example (2)
- Vocabulary for this week
- Group discussion
- Writing activity
- Any questions for today?

Feedback - About the Course

- Goal of course: Allow you to develop skills in statistical reasoning and think critically about the implications and limitations
- Combination of **logical thinking** (lecture, homework, tutorial, final poster project), **mathematics**, **computer simulation** (lecture, homework, poster), and **oral and written discussion/analysis** (homework, tutorial, poster)

Feedback - About the Course

- Communication and discussion is important, will often work in diverse teams
  - Statistician (business environment):
    - May be asked to provide summary of report, answering questions from colleagues who don't have any statistical training
  - Biologist:
    - Statistics used to study biological phenomena
  - Epidemiology:
    - Biostats. used to study risk factors for disease
  - Financial engineers:
    - Combines financial theory, engineering methods, programming and stats tools to solve problems in finance
  - Sports analyst:
    - Stats. used to describe how well players/teams are performing, determine salaries, etc.

Feedback - Course Design

Lecture material (Monday)	- Introduces concepts - Teaches you the skills
Homework questions (During the week)	- Develop hands-on-experience and apply to real-world problems - Learn about analytical issues
Tutorial activities + Final poster project	- Further develop these skills + Communication - Discuss what you observed + Practice ability to explain research - What types of methods did you use, and why? - What did you observe? Did your findings fit your expectations? - Are there any limitations to your findings? - Etc.

* Soon we'll move onto oral presentations in tutorial

Feedback - Communicating your Work

- Short exercises, intended to be no more than half a page, based on homework you should have already completed
- Concepts and questions are discussed during tutorials and in groups
- May be helpful to jot down your ideas first before writing your analysis
  - Analysis should be clear, cohesive, concise, and complete!
  - Have a clear introduction, findings section, and conclusion

Feedback - Communicating your Work

- Analysis should be clear, cohesive, concise, and complete!
- Have a clear introduction, findings section, and conclusion

- Clear: Well-organized, easy to read
- Cohesive: Flows smoothly between parts
- Concise: No wordiness or unnecessary information
- Complete: Covers important parts of project, study, or analysis

Feedback - Communicating your Work

Things to Include in Your Analysis

1. The purpose.
    a. What is it that you're studying? What did you do (methods)? Why should we care about the analytical work you've done?
2. Summary of methods you used.
    a. What did you do? Why did you do it this way?
3. Summary of results.
    a. Include only the most critical things relating to your purpose. Sometimes less is more!
4. Conclusion.
    a. What is your take away message? What did you learn from your analysis?
    b. Conclusion is not a place to present new findings!

Writing Example - Abstract

An Assessment of Oral Health on the Pine Ridge Indian Reservation.
(Gallegos, JR et al.) *Modified for this course*

We assessed the oral health of a group of local Indigenous people living in the Pine Ridge Indian Reservation.
Based on a sample of 292 adults and children, screening personnel counted the number of decayed teeth, total teeth, and dental cavities (both filled and unfilled).

On average, each individual had 4 decayed teeth. Half of adults had 27 or fewer teeth and 26% had an unfilled cavity. Participants had higher numbers of decayed teeth ($p<0.0001$) and lower numbers of filled teeth ($p<0.0001$) than expected.
Amongst the people of Pine Ridge, our study documented a high prevalence of cavities, numerous people with missing teeth, and many unmet dental needs. Future studies of oral health related behaviors, and access to oral health care are needed to explain the dental, periodontal, and soft tissue problems that adversely affect the people of the Pine Ridge Indian Reservation.

Purpose

Methods. Very simple because not statistics-based. You should include more detail here.
Results. Notice how only things critical to their purpose and methods are included. Very concise!
Conclusion (and recommendation)

Writing Example - Abstract

An Assessment of Oral Health on the Pine Ridge Indian Reservation.
(Gallegos, JR et al.) *Modified for this course*

We assessed the oral health of a group of local Indigenous people living in the Pine Ridge Indian Reservation.	*Purpose*
Based on a sample of 292 adults and children, screening personnel counted the number of decayed teeth, total teeth, and dental cavities (both filled and unfilled).	*Methods. Very simple because not statistics-based. You should include more detail here.*
On average, each individual had 4 decayed teeth. Half of adults had 27 or fewer teeth and 26% had an unfilled cavity. Participants had higher numbers of decayed teeth ($p<0.0001$) and lower numbers of filled teeth ($p<0.0001$) than expected.	*Results. Notice how only things critical to their purpose and methods are included. Very concise!*
Amongst the people of Pine Ridge, our study documented a high prevalence of cavities, numerous people with missing teeth, and many unmet dental needs. Future studies of oral health related behaviors, and access to oral health care are needed to explain the dental, periodontal, and soft tissue problems that adversely affect the people of the Pine Ridge Indian Reservation.	*Conclusion (and recommendation)*

Writing Example - Abstract

An Assessment of Oral Health on the Pine Ridge Indian Reservation.
(Gallegos, JR et al.) *Modified for this course*

We assessed the oral health of a group of local Indigenous people living in the Pine Ridge Indian Reservation.	*Purpose*
Based on a sample of 292 adults and children, screening personnel counted the number of decayed teeth, total teeth, and dental cavities (both filled and unfilled).	*Methods. Very simple because not statistics-based. You should include more detail here.*
On average, each individual had 4 decayed teeth. Half of adults had 27 or fewer teeth and 26% had an unfilled cavity. Participants had higher numbers of decayed teeth ($p<0.0001$) and lower numbers of filled teeth ($p<0.0001$) than expected.	*Results. Notice how only things critical to their purpose and methods are included. Very concise!*
Amongst the people of Pine Ridge, our study documented a high prevalence of cavities, numerous people with missing teeth, and many unmet dental needs. Future studies of oral health related behaviors, and access to oral health care are needed to explain the dental, periodontal, and soft tissue problems that adversely affect the people of the Pine Ridge Indian Reservation.	*Conclusion (and recommendation)*

Writing Example - Abstract

An Assessment of Oral Health on the Pine Ridge Indian Reservation.
(Gallegos, JR et al.) *Modified for this course*

We assessed the oral health of a group of local Indigenous people living in the Pine Ridge Indian Reservation.	*Purpose*
Based on a sample of 292 adults and children, screening personnel counted the number of decayed teeth, total teeth, and dental cavities (both filled and unfilled).	*Methods. Very simple because not statistics-based. You should include more detail here.*
On average, each individual had 4 decayed teeth. Half of adults had 27 or fewer teeth and 26% had an unfilled cavity. Participants had higher numbers of decayed teeth ($p<0.0001$) and lower numbers of filled teeth ($p<0.0001$) than expected.	*Results. Notice how only things critical to their purpose and methods are included. Very concise!*
Amongst the people of Pine Ridge, our study documented a high prevalence of cavities, numerous people with missing teeth, and many unmet dental needs. Future studies of oral health related behaviors, and access to oral health care are needed to explain the dental, periodontal, and soft tissue problems that adversely affect the people of the Pine Ridge Indian Reservation.	*Conclusion (and recommendation)*

Example 1

Study skills and students' satisfaction with their performance positively affect their academic achievement. The current research was carried out to investigate the correlation of study skills with academic achievement among the medical and pharmacy students in 2013. This descriptive-analytical study was conducted on 148 students of basic medical sciences and pharmacy through convenience sampling. Data were collected by a valid and reliable questionnaire, consisting of two sections: Demographic information and questions about daily study hours, study skills in six domains, and students' satisfaction with study skills. Collected data sets were analyzed by SPSS-16 software. In total, 10.9% of students were reported to have favorable study skills. The minimum score was found for preparation for examination domain. Also, a significantly positive correlation was observed between students' study skills and their Grade Point Average (GPA) of previous term (P=0.001, r=0.269) and satisfaction with study skills (P=0.001, r=0.493). The findings indicated that students' study skills need to be improved. Given the significant relationship between study skills and GPA, as an index of academic achievement, and satisfaction, it is necessary to promote the students' study skills. These skills are suggested to be reinforced, with more emphasis on weaker domains.

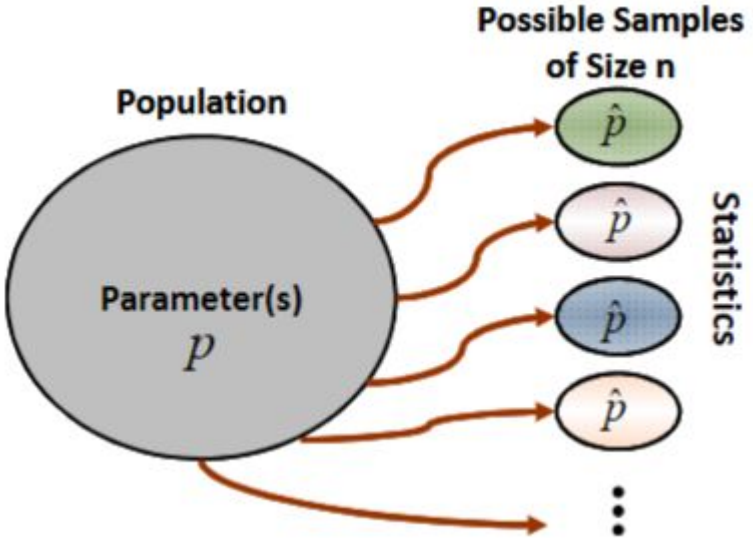
Example 2

This study explored the pattern of video game usage and video game addiction among male college students and examined how video game addiction was related to expectations of college engagement, college grade point average (GPA), and on-campus drug and alcohol violations. Participants were 477 male, first year students at a liberal arts college. In the week before the start of classes, participants were given two surveys: one of expected college engagement, and the second of video game usage, including a measure of video game addiction. Results suggested that video game addiction is (a) negatively correlated with expected college engagement, (b) negatively correlated with college GPA, even when controlling for high school GPA, and (c) negatively correlated with drug and alcohol violations that occurred during the first year in college. Results are discussed in terms of implications for male students' engagement and success in college, and in terms of the construct validity of video game addiction.

Vocabulary for this Week (1/3)

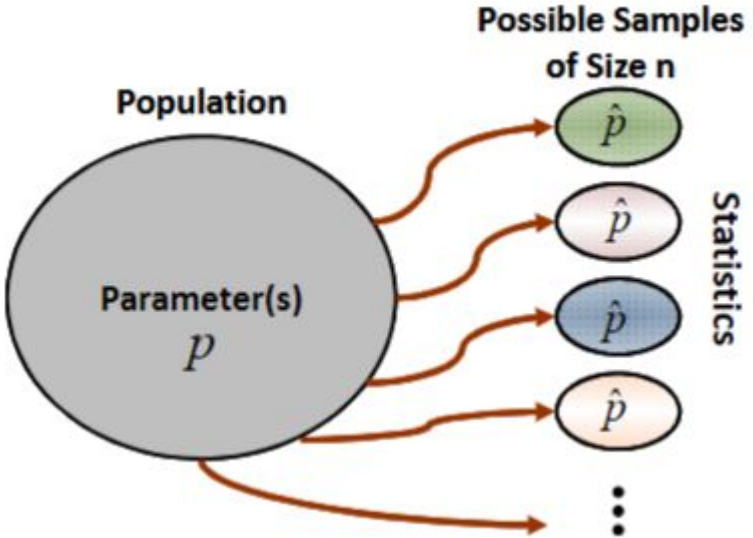
- Sampling distribution
- Population
- Sample
- Parameter
- Statistic

Sampling Distribution



- Sampling distribution of a statistic: Distribution of statistic values taken for all possible samples of the same size (n) from the same population
- Sampling distribution:
  - Take random samples from a certain population
  - Calculate the statistic of each sample
  - Plot and observe distribution

Sampling Distribution (break-down)



- Population: A whole - Every member/element of a group
- Sample: A fraction or percentage of that group
- Parameter (p): Value that tells you something about a population
- Statistic (p̂): Tells you something about a small *part* of the population

https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/what-is-sampling-distribution/v/introduction-to-sampling-distributions

Vocabulary for this Week (2/3)

- Statistical inference
- Significance testing
  - Statistically significant
  - “Due to chance”
- Hypothesis testing
- Null and alternative hypothesis
- Statistically significant
- Meaningful difference
- Significance level
- Simulation
- Strength of evidence

Statistical Inference

- Statistical inference helps make conclusions based on statistical information
  - These conclusions based on statistical inference are uncertain; we're trying to measure the uncertainty
- Significance testing is a type of statistical inference
  - If you calculate or observe something in your data (e.g. a proportion different than expected or a difference between two groups), is it statistically significant OR is it just due to chance?

Conducting a Significance/Hypothesis Test for p

1. State hypotheses
    a. Null hypothesis: No statistical significance
    b. Alternative hypothesis: There is a statistically significant difference
2. Calculate test statistic
    a. Special statistic that helps us decide whether or not the data are compatible with H_0
3. Simulate test statistic assuming H_0 is true
    a. Simulation: Randomly generate samples under the assumption of the null hypothesis
4. Assess evidence against H_0
    a. Use p-values
5. Make a conclusion

How do you interpret a p-value?

- P-value: Probability of obtaining an effect at least as extreme as the one in your sample data, assuming the truth of the null hypothesis
  - Addresses how likely your data is (assuming null holds)
  - Does not measure support for alternative hypothesis

EXAMPLE:

- Suppose that a vaccine effectiveness study produced a p-value of 0.04
- This p-value indicates that if the vaccine had no effect, you'd obtain the observed difference or more in 4% of studies due to random sampling error

What does a p-value NOT mean?

- Statistical significance does not mean practical significance!

EXAMPLE:

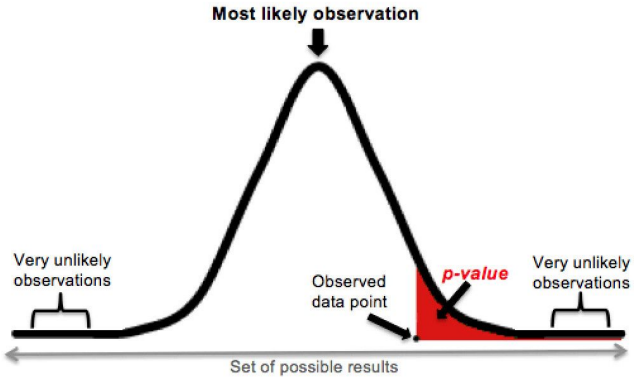
- A clinical trial investigating a new weight loss drug found that people who took their drug lost 0.1 pounds more over the course of a year compared to those who took their competitor's drug (p=0.0001)

Important Things to Note About P-value

- Evidence of statistical significance is either present or it's not
  - Never say something is "almost" statistically significant
- P-value can (almost) never be zero
  - If R gives you zero, it just means you're not looking at enough digits
  - In this case, you should say: $p < 0.0001$!
- You can never accept your null hypothesis!
  - You can only reject (i.e. evidence against null), or fail to reject

Important Things to Note About P-value

Often we assess statistical significance based on a threshold of α=0.05



A *p-value* (shaded red area) is the probability of an observed (or more extreme) result arising by chance

RULE: Reject H_0 if p-value < α

- If p>0.05, then the chance of observing your outcome due to chance alone was greater than 5% (5 times in 100 or more)
- In this case, you would fail to reject the null hypothesis and would not accept the alternative hypothesis

Making a Conclusion

Strength of evidence against H_0

P-value	Evidence
p-value > 0.10	no evidence against H_0
0.05 < p-value < 0.10	weak evidence against H_0
0.01 < p-value < 0.05	moderate evidence against H_0
0.001 < p-value < 0.01	strong evidence against H_0
p-value < 0.001	very strong evidence against H_0

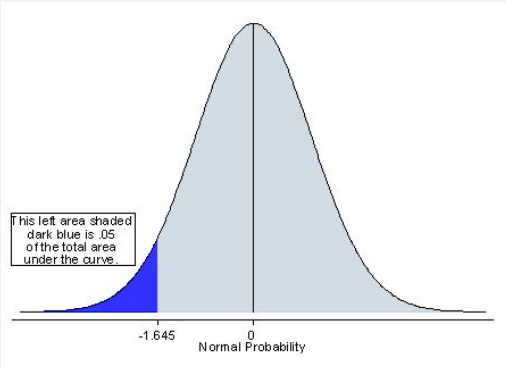
Vocabulary for this Week (3/3)

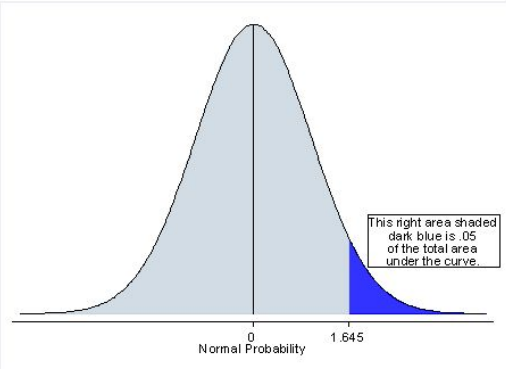
- One-sided test
- Two-sided test

https://piazza.com/class/jpihfcgtfenr9?cid=127

* Assume significance level of 0.05

One-Sided Test





What is it?

- Tests possibility of relationship in one direction

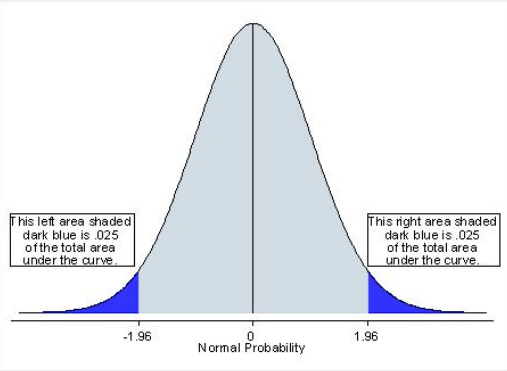
What does it do?

- Tests statistical significance in one direction of interest
- Provides more power to detect an effect
- Consequence: Missing an effect in the other direction
  - If effect in untested direction is negligible, one-sided test can be used

https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/

* Assume significance level of 0.05

Two-Sided Test



What is it?

- Tests possibility of relationship in both directions

What does it do?

- Tests statistical significance in both directions
- E.g. H_0: mean = x
  - Two-sided test will test both if mean is significantly greater than x and if it's significantly less than x

https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/

Relevant R Code from Homework

+ Short explanation

Question (1)

(Adapted from ISRS 3.17) Some people claim that they can tell the difference between Coke and a Pepsi in the first sip. In fact, soda drinkers often have strong preferences for one over the other. A researcher wanting to test the claim that people can tell the difference randomly sampled 50 people. He then filled 50 plain white cups with soda (25 with Coke and 25 with Pepsi) through random assignment, and asked each person to take one sip from their cup and identify the soda as Coke or Pepsi. 32 participants correctly identified the soda brand.

Question (1b) - Based on plot, what is your estimate of the p-value?

(Adapted from ISRS 3.17) Some people claim that they can tell the difference between Coke and a Pepsi in the first sip. In fact, soda drinkers often have strong preferences for one over the other. A researcher wanting to test the claim that people can tell the difference randomly sampled 50 people. He then filled 50 plain white cups with soda (25 with Coke and 25 with Pepsi) through random assignment, and asked each person to take one sip from their cup and identify the soda as Coke or Pepsi. 32 participants correctly identified the soda brand.

```
# distance test statistic is from null value of p
test_stat <- 32/n_observations
distance <- abs(test_stat-0.5)
# to estimate p-value need to find proportion of simulated values at least as far away
# from null value as test statistic
est_pval <- mean(simulated_stats >= 0.5 + distance) + mean(simulated_stats <= 0.5 - distance)
est_pval
```

```
## [1] 0.06
```

Question (2)

A Scottish woman noticed that her husband's smell changed. Six years later he was diagnosed with Parkinson's disease. His wife joined a Parkinson's charity and noticed that odour from other people. She mentioned this to researchers who decided to test her abilities. They recruited 6 people with Parkinson's disease and 6 people without the disease. Each of the recruits wore a t-shirt for a day, and the woman was asked to smell the t-shirts (in random order) and determine which shirts were worn by someone with Parkinson's disease. She was correct for 11 of the 12 t-shirts!

Question (2b) - Simulation to test null hypothesis

```
set.seed(19)  # assume the last 2 digits of my student number are 19
repetitions <- 10000
simulated_stats <- rep(NA, repetitions)
n_observations <- 12
test_stat <- 11/12
for (i in 1:repetitions)
{
  new_sim <- sample(c("correct", "incorrect"), size=n_observations, replace=TRUE)
  sim_p <- sum(new_sim == "correct") / n_observations
  simulated_stats[i] <- sim_p
}
```

```
sim <- data_frame(p_correct = simulated_stats)
ggplot(sim, aes(p_correct)) +
  geom_histogram(binwidth=0.02) +
  geom_vline(xintercept=11/12, color="red") +
  geom_vline(xintercept=0.5-(11/12-.5), color="red")
```

set.seed(n)

Question (2b) - Simulation to test null hypothesis

```
set.seed(19)  # assume the last 2 digits of my student number are 19
repetitions <- 10000
simulated_stats <- rep(NA, repetitions)
n_observations <- 12
test_stat <- 11
for (i in 1:rep
{
  new_sim <- sa
  sim_p <- sum(
  simulated_sta
}
sim <- data_frame(p_correct = simulated_stats)
ggplot(sim, aes(p_correct)) +
  geom_histogram(binwidth=0.02) +
  geom_vline(xintercept=11/12, color="red") +
  geom_vline(xintercept=0.5-(11/12-.5), color="red")
```

set.seed(n):

- **Used so results won't change each time code is run or knitted**
- **Computer saves your random samples to "n"**
  - **Allows you to access the same samples again**

https://piazza.com/class/jpihfcgtfenr9?cid=135

for loop

Question (2b)

for loop:
- Automate repetitions for simulation
- In this example, the code inside the curly brackets ({}) will be evaluated for the number of repetitions we define

```
set.seed(19)  # assume the
repetitions <- 10000
simulated_stats <- rep(NA, repetitions)
n_observations <- 12
test_stat <- 11/12
for (i in 1:repetitions)
{
  new_sim <- sample(c("correct", "incorrect"), size=n_observations, replace=TRUE)
  sim_p <- sum(new_sim == "correct") / n_observations
  simulated_stats[i] <- sim_p
}
sim <- data_frame(p_correct = simulated_stats)
ggplot(sim, aes(p_correct)) +
  geom_histogram(binwidth=0.02) +
  geom_vline(xintercept=11/12, color="red") +
  geom_vline(xintercept=0.5-(11/12-.5), color="red")
```

Question (3)

Coin simulation

Question (3c)

Use 1000 repetitions and set the seed to the last 3 digits of your student number. Note that the probabilities assigned to the values in the vector from which you're sampling using the sample() function are considered equal by default. For example, consider simulating flipping a coin 10 times:

```
sample(c("Head","Tail"),size=10,replace=TRUE)

##  [1] "Tail" "Head" "Head" "Tail" "Head" "Tail" "Head" "Head" "Tail" "Tail"
# will do the same thing as:
sample(c("Head","Tail"),size=10, prob=c(0.5, 0.5), replace=TRUE)

##  [1] "Head" "Head" "Head" "Tail" "Tail" "Head" "Tail" "Tail" "Tail" "Head"
# Even though the exact values are different, if you simulate enough coin flips (by increasing
# the value of 'size', you'll get approximately the same number of "Head" and "Tail" outcomes)

# To modify the code to make Tails much more likely than Heads, we could change the probs:
sample(c("Head","Tail"),size=10,prob=c(0.1, 0.9), replace=TRUE)

##  [1] "Tail" "Tail" "Tail" "Tail" "Tail" "Head" "Head" "Tail" "Tail" "Tail"
```

Question (3c)

Use 1000 repetitions and set the seed to the last 3 digits of your student number. Note that the probabilities assigned to the values in the vector from which you're sampling using the sample() function are considered equal by default. For example, consider simulating flipping a coin 10 times:

```
sample(c("Head","Tail"),size=10,replace=TRUE)
```

```
##  [1] "Tail" "Head" "Head" "Tail" "Head" "Tail" "Head" "Head" "Tail" "Tail"
```

```
# will do the same thing as:
sample(c("Head","Tail"),size=10, prob=c(0.5, 0.5), replace=TRUE)
```

```
##  [1] "Head" "Head" "Head" "Tail" "Tail" "Head" "Tail" "Tail" "Tail" "Head"
```

```
# Even though the exact values are different, if you simulate enough coin flips (by increasing
# the value of 'size', you'll get approximately the same number of "Head" and "Tail" outcomes)

# To modify the code to make Tails much more likely than Heads, we could change the probs:
sample(c("Head","Tail"),size=10,prob=c(0.1, 0.9), replace=TRUE)
```

```
##  [1] "Tail" "Tail" "Tail" "Tail" "Tail" "Head" "Head" "Tail" "Tail" "Tail"
```