Lab 7. Binomial Probability

# Introduction

In this lab we are going to be using binomial probability to determine whether or not heads is a better choice than tails for any particular coin. First, I will walk you through a hypothetical scenario and then you will flip a *real* coin of your choice and test to see whether with that coin your probabilities really are 50-50 for heads or tails.

…honestly this lab really is just about how to find the coin that is the easiest for you to cheat with. Your future self will thank me when you start winning a much larger number of coin tosses!

# Learning Outcomes

By the end of today’s class you should be able to do the following in R:

* Use the dbinom() function to calculate probability
* Use the geom\_bar() function to make a bar plot
* Use the geom\_line() function to make a probability function
* Use the melt() function from the reshape2 package to turn wide-format data into long-format data
* Determine if the coins in your pocket really do have fair odds or not

# Part 1: Fake Coin Flip

## Part 1.1 - Data as Values

To start with, we are going to look at a hypothetical scenario: where I flipped a quarter 10 times and got heads 9 times. I am going to save my results as a variable named **Heads** and a variable named **Tails**.

Heads = 9  
Tails = 1

## Part 1.2 - Data as a Wide Data Frame

I would like to make a bar plot of my data, looking at what I got compared to what I would expect normally with a coin that has 50-50 odds for either side (so, 5 Tails, 5 Heads). To do that I’m going to make a data frame using the data.frame() function. This function will have three columns: **Type** (whether it was expected or what I actually measured), **Heads** (the number of heads expected and measured), and **Tails** (the number of tails expected and measured). Use the c()function to make these columns.

mycoin <- data.frame(Type= c("Expected", "Got"),  
 Heads = c(5, Heads),  
 Tails = c(5, Tails))

## Part 1.3 - Data as a Long Data Frame

What we have just made is called **wide format** data, where we have a lot of columns and very few variables. To plot this using ggplot2 we need to make it into long format data. The easiest way to do that is to use the reshape2 package, which as its name suggests is intended to help you reshape data. You will need to install it first and then load the library.

install.packages("reshape2")  
library("reshape2")

To turn your data from wide format into long format you need to use the melt() function. The results of this function is you have a long format data set that has two new column names: **variable** and **value**.

longcoins <- melt(mycoin)

## Part 1.4 - Data in a Barplot

Now we can use ggplot2 to create a barplot of our data. Don’t forget to load your library!

library("ggplot2")

Use the stat = "identity" and position= "dodge" arguments to separate out your groups nicely.

ggplot(longcoins, aes(x = variable, y = value, fill = Type))+  
 geom\_bar(stat = "identity", position = "dodge")

## Part 1.5 - Exact Probability

In my example, I flipped a coin 10 times and heads came up nine times. Is that actually unusual? Or is that something you could reasonably expect by random chance? To find out, we are going to use the dbinom() function.

This function takes several arguments:

* the number of successes (defined here as the number of heads)
* the total number of trials a.k.a. the number of times I flipped the coin
* the probability of success (or getting heads)

In order, these are **x**, **size**, and **prob**.

dbinom(x = Heads, size = 10, prob = .5)

This gives you the exact probability of getting 9/10 as heads.

### QUESTION 1: What is the probability of getting a single heads out of 10 tosses? Use dbinom() to find your answer

1. 0.001
2. 0.05
3. 0.009
4. 0.1

## Part 1.6 - Probability Distribution Data

Just like any other data, binomial data has a probability distribution. Essentially, given the number of trials that you have, how likely is it to get eight, or seven, or six heads, etc.?

You can look at this using the same function dbinom() with a few modifications. Instead of giving it a single value, we use a colon (:) to specify a range of values - here, from 0 to 10.

dbinom(x=0:10, size=10, prob = .5)

And because we would like to be able to plot this in ggplot2, we will need to make it into a data frame. Here, our x is every possible number of heads from 0 to 10, and our y is the probability for each number of heads.

coin.prob <- data.frame(x = 0:10,   
 y = dbinom(x=0:10, size=10, prob = .5))

## Part 1.7 - Probability Distribution Lineplot

Now we can make a probability distribution, using the geom\_line() function to connect our probabilities and/or using the geom\_bar()function to make a bar plot of those probabilities.

ggplot(coin.prob, aes(x = x, y = y))+  
 geom\_bar(stat="identity")+  
 geom\_line(col = "dodgerblue")

It can be helpful to see where your results lie on this distribution, which is easiest with the geom\_vline() code:

ggplot(coin.prob, aes(x = x, y = y))+  
 geom\_bar(stat="identity")+  
 geom\_line(col = "dodgerblue")+  
 geom\_vline(xintercept = Heads, col = "red")

### QUESTION 2: Look at the probability of getting 9 out of 10 heads. Is this statistically significant? Or should you expect this to happen by random chance a fair amount of the time?

1. p < 0.05, this is something that happens a lot
2. p < 0.05, this is something that is uncommon, and may indicate your coin doesn’t have 50:50 odds.

# Part 2: Double Check Your Understanding

Before we send you off on your own to test the coins in your pocket for maximum cheatability, double check your understanding using the following questions:

### QUESTION 3: If you have 5 heads out of 15 coin tosses, what is the probability of that happening?

1. 0.5
2. 0.05
3. 0.09
4. 0.009

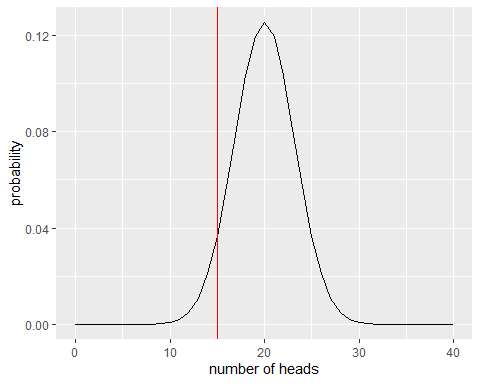
### QUESTION 4: If you have 25 heads out of 150 coin tosses, what is the probability of that happening?

1. 0.5
2. 0.05
3. 0.001
4. < 0.001

### QUESTION 5: Which of the following distributions correctly captures the full probability distribution of 150 coin tosses?

A. ggplot(data.frame(x=1:150, y = dbinom(1:150, 150, prob = .5)), aes(x = x, y = y))+ geom\_line()  
  
B. ggplot(data.frame(x=0:150, y = dbinom(0:150, 150, prob = .5)), aes(x = x, y = y))+ geom\_line()  
  
C. ggplot(data.frame(x=0:15, y = dbinom(0:15, 150, prob = .5)), aes(x = x, y = y))+ geom\_line()

### QUESTION 6: According to the graphic below, where the black line is the probability distribution for that number of trials and the red line is the number of heads actually counted, is this number of heads commonly expected by random chance? (Hint: look at where the red line and black line cross on the y axis!)



1. Yes, this number of heads happens a fair amount
2. No, this number of heads is not very likely (p < 0.05)
3. No, this number of heads is not likely at all (p < 0.001)

# Part 3: How to Cheat and Prosper

In this section you are going to experiment with a physical coin and create an R script walkthrough of your results. The goal is to see if your coin really is behaving according to a 50:50 probability - and if it isn’t, you should hold onto that and use it to turn the odds in your favor in the future, muahahaha.

For all of the answers below, heads are considered a “positive” result.

## Part 3.1 - Your Coin

For this part of the lab, you are going to flip a real coin. I know, real objects, so exciting!

Ahem.

Anyways, find a real coin of some kind. It can be whatever you like - e.g. a dirty penny from 1995, a well-cared for Susan B Anthony coin, a shiny new Euro, a coin from an Arcade that shut down in the 80’s, etc.

You now need to describe what this coin is for us, and also hypothesize why it might matter that we know what type of coin you are using.

### QUESTION 7: What type of coin are you using for your experiment?

### QUESTION 8: Why do you think the type, age, and condition of the coin matters? How might that affect probability?

For all the following, make sure to record your steps in an R script. We have supplied you with a blank version (**Lab 7 Part 3 Blank Script.R**) or you can make your own file.

## Part 3.2 - Flip your coin 30 times

Record the number of heads and tails you get, recording the total number of each as values in your R script similar to part 1.1. These should add up to 30, or you’re going to have some problems!

## Part 3.3 - Find the exact probability of your results

Use part 1.5 to help you find the exact probability of your specific number of heads. Copy your code onto the R script.

## Part 3.4 - Create a probability plot of 30 coin tosses.

Edit the dbinom() code and geom\_line code from part 1.6-1.7 so they match your data and place these in your script.

### QUESTION 9: Upload your script for the above sections. Copy-paste your code, do not upload a picture.

### QUESTION 10: According to your data, do you suspect your coin has 50:50 odds? Use your *p*-value to support your answer.

### QUESTION 11: Look at the probability plot of 30 coin tosses. How many heads would you have to get in order to be suspicious that your coin did not have 50:50 odds of getting a heads (or positive) flip result?

# Part 4 - Interpretation

Read through this Smithsonian Magazine article on coin flipping percentages: <https://www.smithsonianmag.com/science-nature/gamblers-take-note-the-odds-in-a-coin-flip-arent-quite-5050-145465423/>

The long and short of it is that it turns out that many coins don’t actually have 50:50 odds of landing on one side or the other (all of your childhood soccer referees LIED to you!). The odds might not be THAT different, but 51:49 is still a difference. Your coin may or may not have similar issues.

### QUESTION 12: How might you change your dbinom() code if you knew your coin actually didn’t have a 50/50 shot of landing on one side? What argument would you need to modify?

THE END!