

Lesson 2 MA206Y

Nicholas Clark

This lesson introduces us to some very important concepts in the study of statistics. Namely a **sample**, a **statistic**, **sample size**, and a **parameter**. I like to think of this in the following way:

In order to actually do anything, we assume that our random process follows some **model**. You may have heard the axiom, “All models are wrong but some are useful”, in the study of statistics we really take this to heart. We know that even if our model accurately reflects our mechanism of interest, nature likely gets a vote, so there’s a randomness in what we observe. For instance, if we have a coin, we might assume that $P(\text{heads}) = p$ and $P(\text{tails}) = 1 - p$. This is an example of a statistical model. In order to see if our coin is fair, perhaps we flip it four times. Our sample size then is:

Our sample could be:

Our statistic might be:

And our parameter is

If our coin is fair our parameter is:

In this case we could write out all of our possible samples, this is sometimes called the sample space:

If our statistic is the number of heads each of our samples has a value, let’s write them:

If our coin was fair, how rare would it be for us to observe four heads? Can we quantify this?

If you had to wager \$1000 that this is a fair coin based off of this experiment would you do it? Why or why not?

How would we strengthen our argument?

One of the great things about R is that it allows us to run experiments cheaply. If $p = 0.5$ we can run our experiment using the following code:

```
library(tidyverse)
possible.outcomes=2 #Either we get a heads or a tails
p=.5 #This assumes our coin is fair
sample.size=4

sample=rbinom(sample.size,possible.outcomes-1,p)
sample
statistic=sum(sample)
```

Of course this is just running our experiment once, let's run our experiment 1000 times!

```
possible.outcomes=2 #Either we get a heads or a tails
p=.5 #This assumes our coin is fair
sample.size=4
num.experiments=1000
all.of.my.stats=data.frame(trial=seq(1,num.experiments),stats=NA) #I am making a blank object that I'm
for(j in 1:num.experiments){
  sample=rbinom(sample.size,possible.outcomes-1,p)
  all.of.my.stats[j,]$stats=sum(sample)
}
```

So now we can look at this:

```
all.of.my.stats %>% ggplot(aes(x=stats))+geom_histogram()
```

So how rare is it that we observe four heads?

```
all.of.my.stats %>% filter(stats==4)%>%summarise(total=n(),perc.total=n()/num.experiments)
```

What if we wanted to do flip our coin 28 times?

```
possible.outcomes=2 #Either we get a heads or a tails
p=.5 #This assumes our coin is fair
sample.size=28
num.experiments=1000
all.of.my.stats=data.frame(trial=seq(1,num.experiments),stats=NA) #I am making a blank object that I'm
```

```
for(j in 1:num.experiments){  
  sample=rbinom(sample.size,possible.outcomes-1,p)  
  all.of.my.stats[j,]$stats=sum(sample)  
}
```

In the dolphin study, if the dolphin was just guessing how rare would it be to observe 16 successes? How does this relate to our heads/tails experiment?

Three S's. Statistic, simulate, quantify strength of evidence:

1-17 From “Can Dogs Understand Human Cues”. Due next lesson, 25 instructor points.