5.3: Is there a difference?

Let's get back to experiments. In an experiment we want to know if our independent variable (our manipulation) causes a change in our dependent variable (measurement). If this occurs, then we will expect to see some differences in our measurement as a function of manipulation.
Consider the light switch example:

Light Switch Experiment: You manipulate the switch up (condition 1 of independent variable), light goes on (measurement). You manipulate the switch down (condition 2 of independent variable), light goes off (another measurement). The measurement (light) changes (goes off and on) as a function of the manipulation (moving switch up or down).
You can see the change in measurement between the conditions, it is as obvious as night and day. So, when you conduct a manipulation, and can see the difference (change) in your measure, you can be pretty confident that your manipulation is causing the change.

> **Note**
>
> To be cautious we can say "something" about your manipulation is causing the change, it might not be what you think it is if your manipulation is very complicated and involves lots of moving parts

---

Chance can produce differences

Do you think random chance can produce the appearance of differences, even when there really aren't any? I hope so. We have already shown that the process of sampling numbers from a distribution is a chancy process that produces different samples. Different samples are different, so yes, chance can produce differences. This can muck up our interpretation of experiments.
Let's conduct a fictitious experiment where we expect to find no differences, because we will manipulate something that shouldn't do anything. Here's the set-up:
You are the experimenter standing in front of a gumball machine. It is very big, has thousands of gumballs. 50% of the gumballs are green, and 50% are red. You want to find out if picking gumballs with your right hand vs. your left hand will cause you to pick more green gumballs. Plus, you will be blindfolded the entire time. The independent variable is Hand: right hand vs. left hand. The dependent variable is the measurement of the color of each gumball.
You run the experiment as follows. 1) put on blind fold. 2) pick 10 gumballs randomly with left hand, set them aside. 3) pick 10 gumballs randomly with right hand, set them aside. 4) count the number of green and red gumballs chosen by your left hand, and count the number of green and red gumballs chosen by your right hand. Hopefully you will agree that your hands will not be able to tell the difference between the gumballs. If you don't agree, we will further stipulate the gumballs are completely identical in every way except their color, so it would be impossible to tell them apart using your hands. So, what should happen in this experiment?
"Umm, maybe you get 5 red gum balls and 5 green balls from your left hand, and also from your right hand?". Sort of yes, this is what you would usually get. But, it is not all that you can get. Here is some data showing what happened from one pretend experiment:

```
hand<-rep(c("left","right"),each=10)
gumball<-rbinom(20,1,.5)
df<-data.frame(hand,gumball)
knitr::kable(df)
```

run | restart | restart & run all

```
|hand  | gumball|
|:-----|-------:|
|left  |       1|
|left  |       1|
|left  |       1|
|left  |       1|
```

```
|left  |       0|
|left  |       0|
|left  |       0|
|left  |       0|
|left  |       0|
|left  |       0|
|right |       0|
|right |       0|
|right |       0|
|right |       0|
|right |       1|
|right |       0|
|right |       0|
|right |       0|
|right |       1|
|right |       1|
```

“What am I looking at here”. This is a long-format table. Each row is one gumball. The first column tells you what hand was used. The second column tells you what kind of gumball. We will say 1s stand for green gum balls, and 0s stand for red gumballs. So, did your left hand cause you to pick more green gumballs than your right hand?

It would be easier to look at the data using a bar graph. To keep things simple, we will only count green gumballs (the other gumballs must be red). So, all we need to do is sum up the 1s. The 0s won’t add anything.

```
library(ggplot2)
hand<-rep(c("left","right"),each=10)
gumball<-rbinom(20,1,.5)
df<-data.frame(hand,gumball)
sum_df<-aggregate(gumball~hand,df,sum)
ggplot(sum_df,aes(x=hand,y=gumball))+
  geom_bar(stat="identity")+
  theme_classic()
```

run | restart | restart & run all

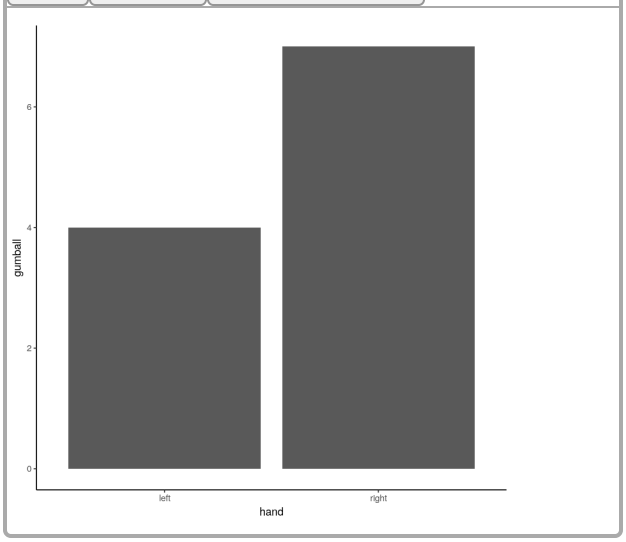

Figure \(\PageIndex{1}\): Counts of gumballs picked.

Oh look, the bars are not the same. One hand picked more green gum balls than the other. Does this mean that one of your hands secretly knows how to find green gumballs? No, it's just another case of sampling error, that thing we call luck or chance. The difference here is caused by chance, not by the manipulation (which hand you use). Major problem for inference alert. We run experiments to look for differences so we can make inferences about whether our manipulations cause change in our measures. Now we know that we can find differences by chance. How can we know if a difference is real, or just caused by chance?

Differences due to chance can be simulated

Remember when we showed that chance can produce correlations. We also showed that chance is restricted in its ability to produce correlations. For example, chance more often produces weak correlations than strong correlations. Remember the window of chance? We found out before that correlations falling outside the window of chance were very unlikely. We can do the same thing for differences. Let's find out just what chance can do in our experiment. Once we know what chance is capable of we will be in a better position to judge whether our manipulation caused a difference, or whether it could have been chance.

The first thing to do is pretend you conduct the gumball experiment 10 times in a row. This will produce 10 different sets of results. For each of them we can make a bar graph, and look at whether the left hand chose more green gumballs than red gumballs. It looks like this:

```
library(ggplot2)
hand<-rep(rep(c("left","right"),each=10),10)
experiment<-rep(1:10,each=20)
gumball<-rbinom(20*10,1,.5)
df<-data.frame(experiment,hand,gumball)
sum_df<-aggregate(gumball~experiment*hand,df,sum)
ggplot(sum_df,aes(x=hand,y=gumball))+
  geom_bar(stat="identity")+
  theme_classic()+
  facet_wrap(~experiment)
```

run | restart | restart & run all

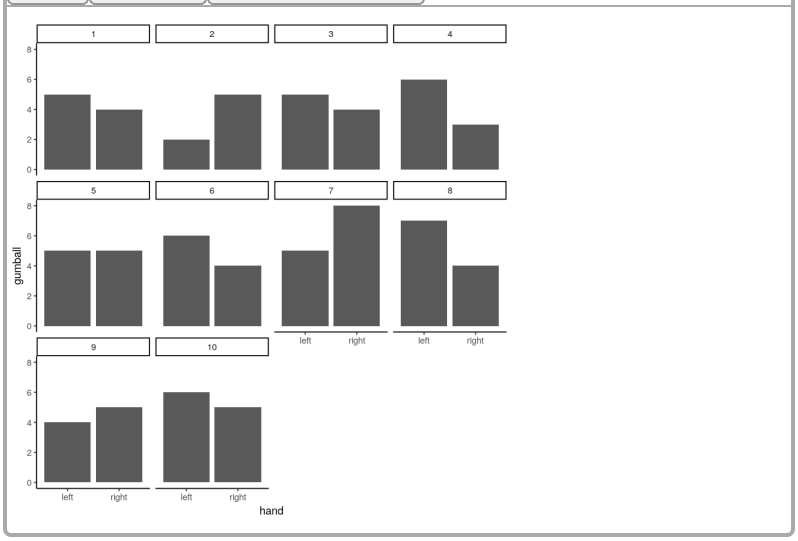

Figure \(\PageIndex{2}\): 10 simulated replications of picking gumballs. Each replication gives a slightly different answer. Any difference are all due to chance, or sampling error. This shows that chance alone can produce differences, just by the act of sampling.

These 10 experiments give us a better look at what chance can do. It should also mesh well with your expectations. If everything is left up to chance (as we have made it so), then sometimes your left hand will choose more green balls, sometimes your right hand will choose more green gumballs, and sometimes they will choose the same amount of gumballs. Right? Right.