Lesson 6 MA206Y

Nicholas Clark

Admin

A study in *Psychonomic Bulletin and Review* presented evidence that "people use facial prototypes when they encounter different names."

Who is Tim and Who is Bob?

The parameter of interest here is the probability that a student in your class would assign the same name to the face on the left.

What is H_0 and H_a?

What is $\hat{p}$? i.e. what proportion of the class put Tim's name on the left?

We can also use a standardized statistic:

z= (statistic - mean of null distribution)/sd(null distribution)

In R:

```
possible.outcomes=2 #Either we get a heads or a tails
p=.5 #This assumes our coin is fair
sample.size=18 #I'm flipping my coin 4 times2
num.experiments=1000 #I have 1000 students doing the experiment
all.of.my.stats=data.frame(trial=seq(1,num.experiments),stats=NA)
#I am making a blank object that I'm going to fill in
for(j in 1:num.experiments){
  sample=rbinom(sample.size,possible.outcomes-1,p)
  all.of.my.stats[j,]$stats=sum(sample)/18
}
mean.of.null=mean(all.of.my.stats$stats)
sd.of.null = sd(all.of.my.stats$stats)
```

So our z statistic is:

According to our book, what does this tell us?

According to page 49, what else does this value tell us?

What if we observed the opposite of what we observed? What would $\hat{p}$ be?

Would this change our conclusion?

Let's say the ratio stayed the same but we had 180 students in the class? What would happen to our z value?

Perhaps we want to say something stronger than what is in our text. As it turns out, as long as certain *validity conditions* are met we know the distribution of our z statistic, and we can calculate it without doing a single simulation.

If we have at least 10 successes and at least 10 failures, we can assume $Z = \frac{\hat{p}-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$ has a standard normal (or bell shaped) distribution.

The standard normal distribution looks like:
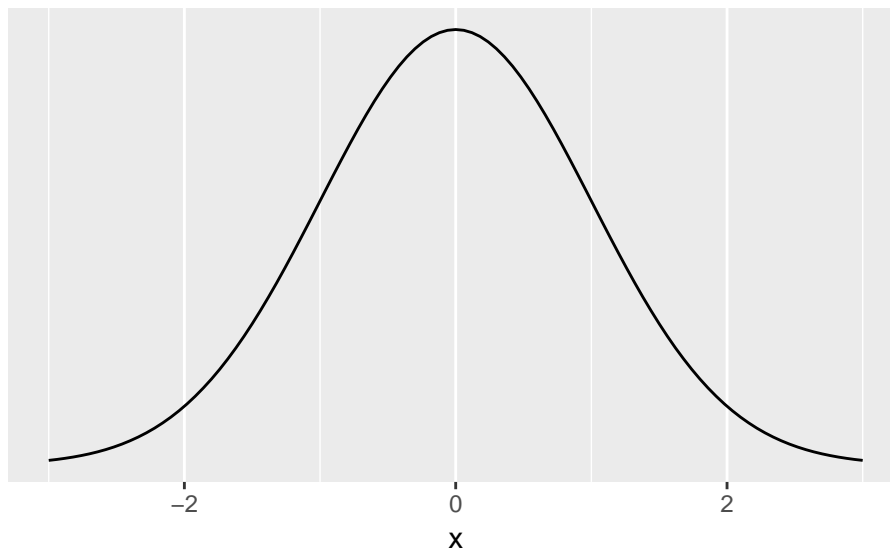
Knowing this is extremely convenient as we no longer have to simulate to find probabilities. Say for instance our $Z = -1.3$ and we want to know how rare it would be, under H_0 to observe something as small or smaller than $Z = -1.3$, we could simply do:

```
pnorm(-1.3)
```

```
## [1] 0.09680048
```

What the

```
pnorm(-1.3)
```

command does is it integrates the above curve from ∞ to -1.3

Going back to our simulations we were doing before this would be the same as:

```
possible.outcomes=2 #Either we get a heads or a tails
p=.5 #This assumes our coin is fair
sample.size=400 #I'm flipping my coin 400 times so validity conditions definitely met
num.experiments=1000 #I have 1000 students doing the experiment
all.of.my.stats=data.frame(trial=seq(1,num.experiments),stats=NA)
#I am making a blank object that I'm going to fill in
```

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

all.of.my.stats %>% ggplot(aes(x=stats))+geom_histogram(aes(y=..density..),breaks=seq(-4, 4, by=.2))
```
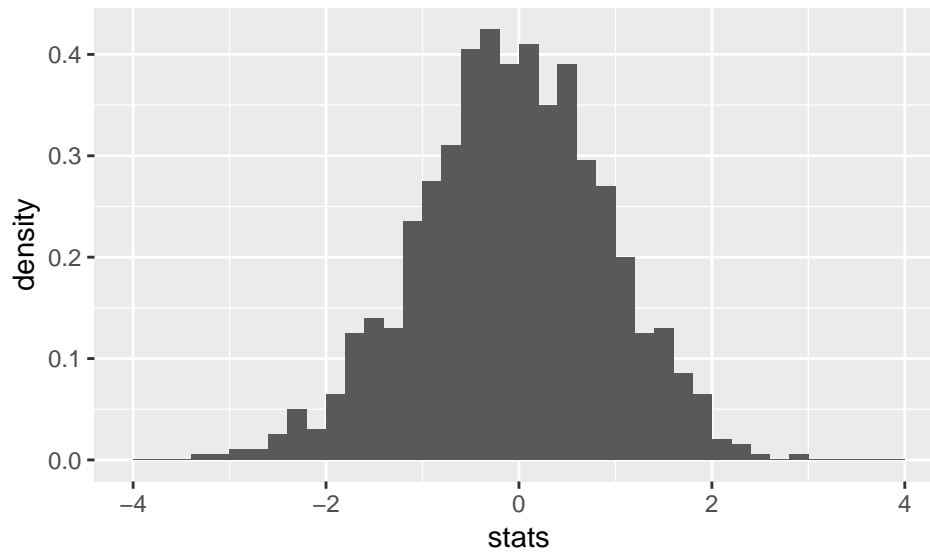

Which we can approximate by the normal distribution:
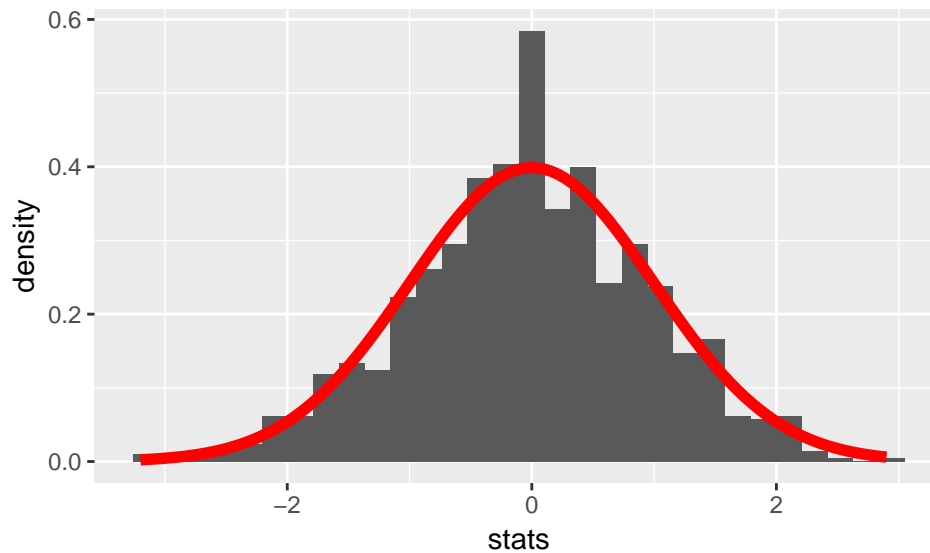
For our experiment are the validity conditions met?

Add 10 to both the number of lefts and the number of rights, using simulation test the null and alternative hypothesis:

$$H_0 : \pi = 0.5 \quad H_a : \pi \neq 0.5$$

What is $\hat{p}$ in our experiment?

What is Z in our experiment?

Using simulation, how many Z sims are both higher than the positive of our Z and lower than the negative of our Z. Note here we have a 'two-sided' hypothesis so more extreme means both higher or lower than what we observed.

Are our validity conditions met?

Repeat the analysis but now use the theory-based approach. Note pnorm(Z) gives you everything to the **left** of Z, so if we want everything to the **right** of Z we would need to do 1-pnorm(Z) since the total area under the curve is 1.