Lsn 17 - MA206Y

Clark

Admin

Problem 5.3.7 (a-d, g, h)

Determine whether hte type of bicycle frame affects travel time to work. Dr. Groves tossed a coin to decide whether he would ride to work on a carbon frame or a steel frame and recorded the weight for 56 days.

Our Null hypothesis and alternative hypothesis are:

Our response variable is:

Our parameter of interest is now:

Note that it would NOT be appropriate to use π_{steel} and π_{carbon} here. Why?

One statistic that we could use to explore our parameter of interest is:

After we have decided what statistic to use we do our experiment and find the realization of our test statistic, which in this case is:

Recall that in order to find out how rare our test statistic would be, we need to know the distribution of all possible $\bar{X}_1 - \bar{X}_2$ under H_0.

```r
bike.dat<-read.table("http://www.isi-stats.com/isi/data/chap6/BikeTimes.txt",header=T)
group.means<-bike.dat %>% group_by(Frame)%>%summarize(xbar=mean(Time))
our.stat<-group.means$xbar[1]-group.means$xbar[2] #Sorry!  I don't know a more straight forward way
#to code this!
```

Note that if H_0 is true, the labeling doesn't matter. So we can shuffle the labels around:

```r
bike.dat.shuff<-bike.dat %>% mutate(shuff.labels=sample(Frame))
```

And under that shuffle we can see what we would get for a statistic

```r
group.means<-bike.dat.shuff %>% group_by(shuff.labels)%>%summarize(xbar=mean(Time))
H0.stat<-group.means$xbar[1]-group.means$xbar[2]
```

So we can do this 1000 times to get a feel for the distribution under H_0

```r
results<-data.frame(repetition=seq(1,1000),stat=NA)
for(j in 1:1000){
  bike.dat.shuff<-bike.dat %>% mutate(shuff.labels=sample(Frame))
  group.means<-bike.dat.shuff %>% group_by(shuff.labels)%>%summarize(xbar=mean(Time))
  H0.stat<-group.means$xbar[1]-group.means$xbar[2]
  results[j,]$stat=H0.stat
}
```

So our simulation distribution is:

```r
results %>% ggplot(aes(x=stat))+geom_histogram()+
  geom_vline(xintercept=our.stat,color="red")+
  geom_vline(xintercept=-our.stat,color="red")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
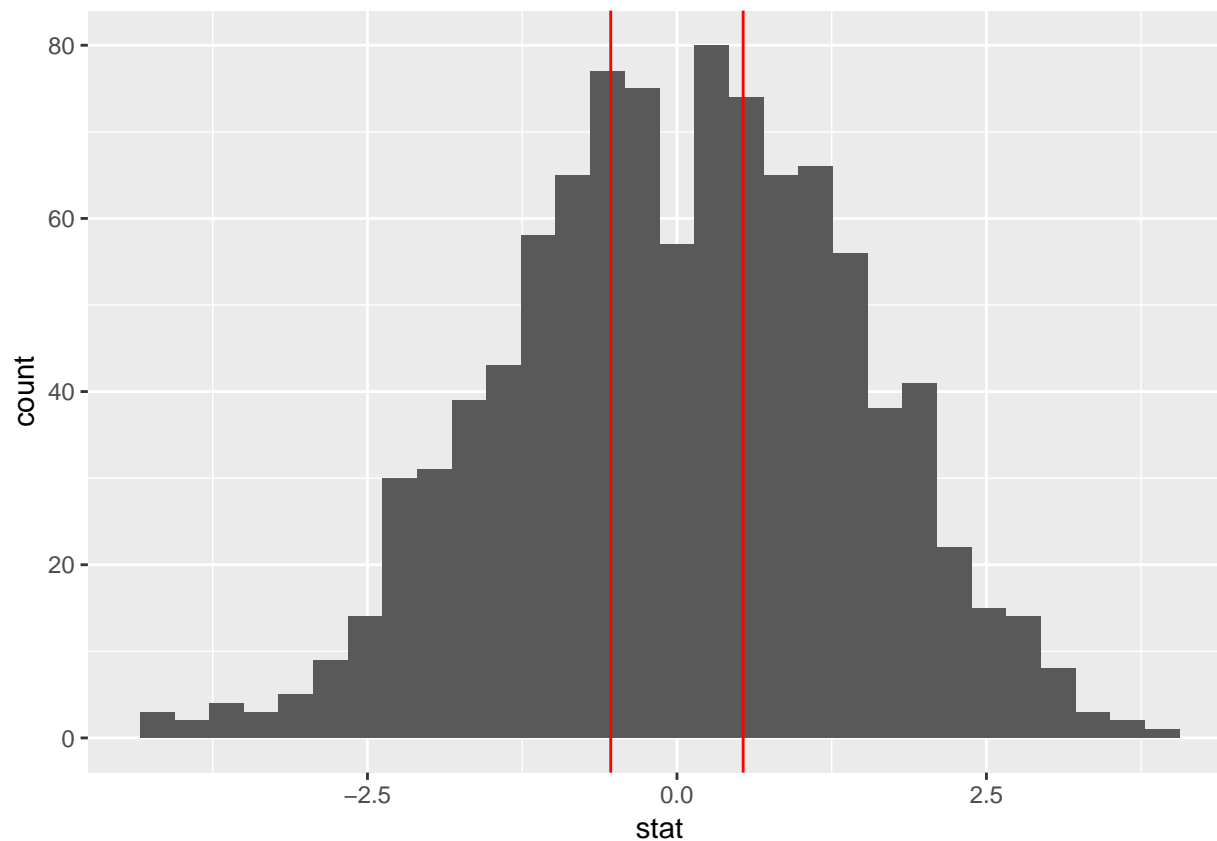
So, if H_0 is true, would it be rare to observe what we actually observed?

Why is our distribution under H_0 centered at 0?

The same thing can be done with the applet.

Note in either case we can use the 2SD method to construct an approximate 95% CI.

```
sd(results$stat)
```

```
## [1] 1.405223
```

Compare to applet

Practice

Let's work through 6.2.24 on the boards.