Lsn 13 - MA206Y

Clark

Admin

In 1979 there was a study conducted on the effect of cigarette smoking on the amount of air an individual can exhale in the first second of a foreceful breath.

The data collected were age, fev, heigh, sex, and smoke.

According to the research question, what would you consider the explanatory variable and what would you consider the response variable?

What are our observational units?

Is our response variable continuous or categorical?

```
#install.packages("isdals")
library(isdals)
data(fev) #Another way to call in data is to use data that are tied to a library
fev %>% ggplot(aes(x=FEV))+geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Our data are coded 0 if not smoke, 1 if smoke. We run the code:

```
fev %>% group_by(Smoke)%>%summarize(mean.FEV=mean(FEV))
```

```
## # A tibble: 2 x 2
##   Smoke mean.FEV
##   <int>    <dbl>
## 1     0     2.57
## 2     1     3.28
```

What is the $\bar{X}_{\text{Smoker}}$ and what is $\bar{X}_{\text{Non Smoker}}$?

Is this surprising?

Let's draw our structure or causal diagram:

Can we conclude that smoking causes a higher FEV? Why or why not?

What could be a confounding variable here?

Let's draw another causal diagram:

```
fev %>% ggplot(aes(x=jitter(Age),y=FEV,color=as.factor(Smoke)))+geom_point()
```

What do we see here? How is this potentially evidence of confounding?

```
fev %>% group_by(Smoke)%>%summarize(mean.Age=mean(Age))
```

```
## # A tibble: 2 x 2
##   Smoke mean.Age
##   <int>    <dbl>
## 1     0     9.53
## 2     1    13.5
```

Is this an experiment or an observational study?

How would we make this an experiment? Is this ethical?

It's very difficult (though not impossible) to draw cause and effect relationships in observational studies. In an observational study, the values of the explanatory variable are simply observed. Here our explanatory variable was smoking. In an experiment, we get ot assign the conditions to the different subjects.

An area of research in biomechanics and gerontology concerns falls and fall-related injuries, especially for elderly peple. Recent studies have focused on how individuals respond to large postural disturbances. One question is whether subjects can be instructed to improve their recovery from such disturbances. Suppose researchers want to compare two such recovery strategies, lowering and elevating. Subjects will have first been

trained on one of these two recovery strategies, and they will be asked to apply it after they feel themselves tripping. The researchers will then induce the subject to trip while walking using a concealed mechanical obstacal. Suppose we have 24 subjects, 8 females and 16 males.

One way to design this study would be to assign the 8 females to use the elevating strategy and the 16 males to use the lowering strategy. Is this a reasonable experiment?

Another way is to assign 4 females and 8 males to each group. Clearly the proportions are the same, but what are we trying to achieve by using this strategy?

Now, if you see a difference in the proporiton of trips between the two groups, could it be because of the sex of the subject? Could it be due to other variables distinct from the recovery strategy?

How else could we conduct the study?

Let's say we decide to use a completely randomized design (CRD). Let's go to the Randomizing Subjects applet (www.rossmanchance.com/ISIapplets.html). Do the randomization and look at the proportion of males and females in each group are we completely balanced between Males and Females? Click randomize again, what happens?

Let's randomize 1000 times (turn of animate). On *average* what can we say about the difference in sample proportions?

Do CRDs always balance possible confounding variables? Is there a tendency for there to be a similar proportion of men in the two groups?

Instead we choose to employ a randomized block design (RBD). Click on the randomized block. Now observe the sex proportion. Are we always balanced?

As it turns out, height also might be a confounder. Under the CRD design are we, on average, balanced with respect to height? What about under the RBD design? On average is the RBD or the CRD more variable with respect to the differnce in height?

Is this surprising?

What if there's some gene some folks have. On average using the CRD are we balanced with respect to that gene? What about the RBD? What we should find is that blocking by gender helps to control for confounders that we are blocking on and other confounders that are associated with gender, but does little to help control for other confounders that might be present. We can never eliminate all variability in our model, what we can hope to do is distribute our unexplained variability (on average) equally across all of our treatment groups.

Chart on pg 244

Midterm Review