Lesson 22

Clark

Admin

Brain size makes you smarter...

```
brain.dat<-read.table("http://www.isi-stats.com/isi2/data/BrainSize.txt",header=T)
brain<-brain.dat %>% select(PIQ,Height,BrainSize)
```

The first question asked is how does Brain size impact PIQ ignoring all other possible covariates. The model is:

We fit this as:

```
brain.lm<-lm(PIQ~BrainSize,data=brain)
summary(brain.lm)
```

```
##
## Call:
## lm(formula = PIQ ~ BrainSize, data = brain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.079 -17.508  -2.096  17.100  41.574
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.66036   43.71288   0.107   0.9157
## BrainSize    0.11765    0.04806   2.448   0.0194 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.21 on 36 degrees of freedom
## Multiple R-squared:  0.1427, Adjusted R-squared:  0.1189
## F-statistic: 5.993 on 1 and 36 DF,  p-value: 0.01937
```

If we look at the plot, it looks pretty good:

```
brain %>% ggplot(aes(x=BrainSize,y=PIQ))+geom_point()+
  stat_smooth(method=lm,se=F)
```

But can we say that this relationship is causal?

```
ggpairs(brain)
```

Here we see a couple of things

In order to account for the potential covariation we want to 'adjust' for height. That is, we only want to consider the variation in the model that can be isolated from the variation due to height.

Note that height adjusted PIQ is found from:

```
brain %>% ggplot(aes(x=BrainSize,y=PIQ,color=Height))+geom_point()+
    scale_color_gradient(low="blue", high="red")
```

Now we can adjust by examining the residuals of the univariate model

```
uni.lm<-lm(PIQ~Height,data=brain)
brain<-brain %>% mutate(adj.PIQ=uni.lm$residuals)
adj.lm<-lm(adj.PIQ~BrainSize,data=brain)
coef(adj.lm)
```

```
## (Intercept)    BrainSize
## -122.159719     0.134722
```

What happened to the slope? Could we have predicted this?

Note that we still haven't adjusted Brain Size for height. To expand on this, let's draw a causal diagram:

To do this, we want to height adjusted brain size. TO get this we need the residuals from the model:

```
brain.adj.lm<-lm(BrainSize~Height,data=brain)
brain<-brain %>% mutate(adj.Brain=brain.adj.lm$residuals)
```

Now we fit the model:

```
adj.lm<-lm(adj.PIQ~adj.Brain,data=brain)
coef(adj.lm)
```

```
## (Intercept)     adj.Brain
## 8.555482e-17 2.060561e-01
```

In words, we are fitting:

That is, if we account for height, then change in brain size by one unit predicts a change in .2061 units for PIQ, whereas previously it was 0.117. What is going on here?

Note here, if height is NOT related to Brain size, then the model

Would yield residuals that look just like our origial values.

To see if both variables explain a sufficient amount of variation in our response we run:

```
library(car)
full.lm<-lm(PIQ~BrainSize+Height,data=brain)

Anova(full.lm,type="II")
```

```
## Anova Table (Type II tests)
##
## Response: PIQ
##            Sum Sq Df F value     Pr(>F)
## BrainSize  5408.1  1 14.2077 0.0006051 ***
## Height     2875.4  1  7.5541 0.0094035 **
```

```
## Residuals 13322.5 35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

NOte here that the total Sums of squares are:

```
SST=sum((brain$PIQ-mean(brain$PIQ))^2)
```

Let's look at the ANOVA table, what do we see that's surprising?

What is going on here?

Our fitted model is:

```
coef(full.lm)
```

```
## (Intercept)    BrainSize       Height
## 111.2784739   0.2060561   -2.7298407
```

Of course we could add an interaction term too. The model is:

What is the interpretation of the interaction term?

```
inter.lm<-lm(PIQ~BrainSize*Height,data=brain)
summary(inter.lm)
```

```
##
## Call:
## lm(formula = PIQ ~ BrainSize * Height, data = brain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.779 -12.001  -3.871  14.209  51.604
##
## Coefficients:
```

```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.295e+01  6.922e+02   0.134    0.894
## BrainSize        2.260e-01  7.516e-01   0.301    0.766
## Height          -2.461e+00  1.017e+01  -0.242    0.810
## BrainSize:Height -2.914e-04  1.097e-02  -0.027    0.979
##
## Residual standard error: 19.79 on 34 degrees of freedom
## Multiple R-squared:  0.2949, Adjusted R-squared:  0.2327
## F-statistic:  4.74 on 3 and 34 DF,  p-value: 0.007219
```

whoa...? What is going on?

```
brain2<-brain.dat %>% select(PIQ,Height,BrainSize,interaction)
ggpairs(brain2)
```

So, what is going on is that the ineraction term is highly correlated with Brain Size. SO, let's think through the interpretation β terms.

When you're in this situation one technique is to standardize the variables

```
brain.cov<-brain%>%mutate_at(funs(scale(.)),.vars=vars(-PIQ))
std.lm<-lm(PIQ~Height*BrainSize,data=brain.cov)
summary(std.lm)
```

```
##
## Call:
## lm(formula = PIQ ~ Height * BrainSize, data = brain.cov)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.779 -12.001  -3.871  14.209  51.604
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     111.39048    3.69138  30.176  < 2e-16 ***
## Height          -10.88380    4.08511  -2.664 0.011712 *
## BrainSize        14.95030    4.02498   3.714 0.000728 ***
## Height:BrainSize  -0.08444    3.17809  -0.027 0.978958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.79 on 34 degrees of freedom
## Multiple R-squared:  0.2949, Adjusted R-squared:  0.2327
## F-statistic:  4.74 on 3 and 34 DF,  p-value: 0.007219
```