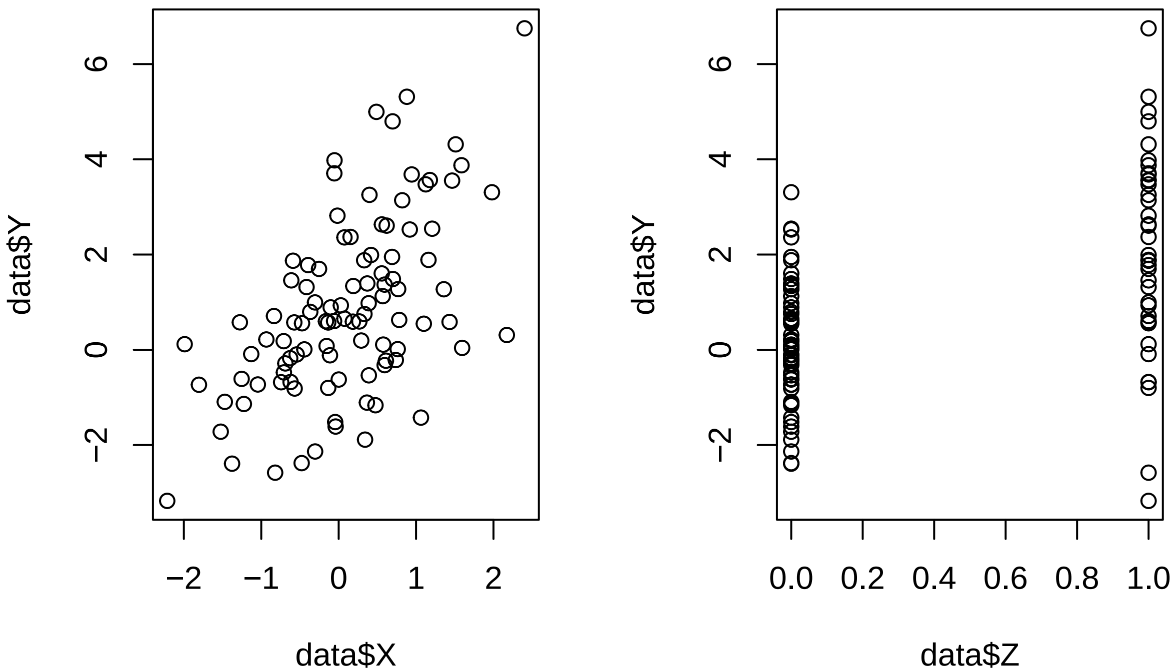
Today

Nicholas Ray

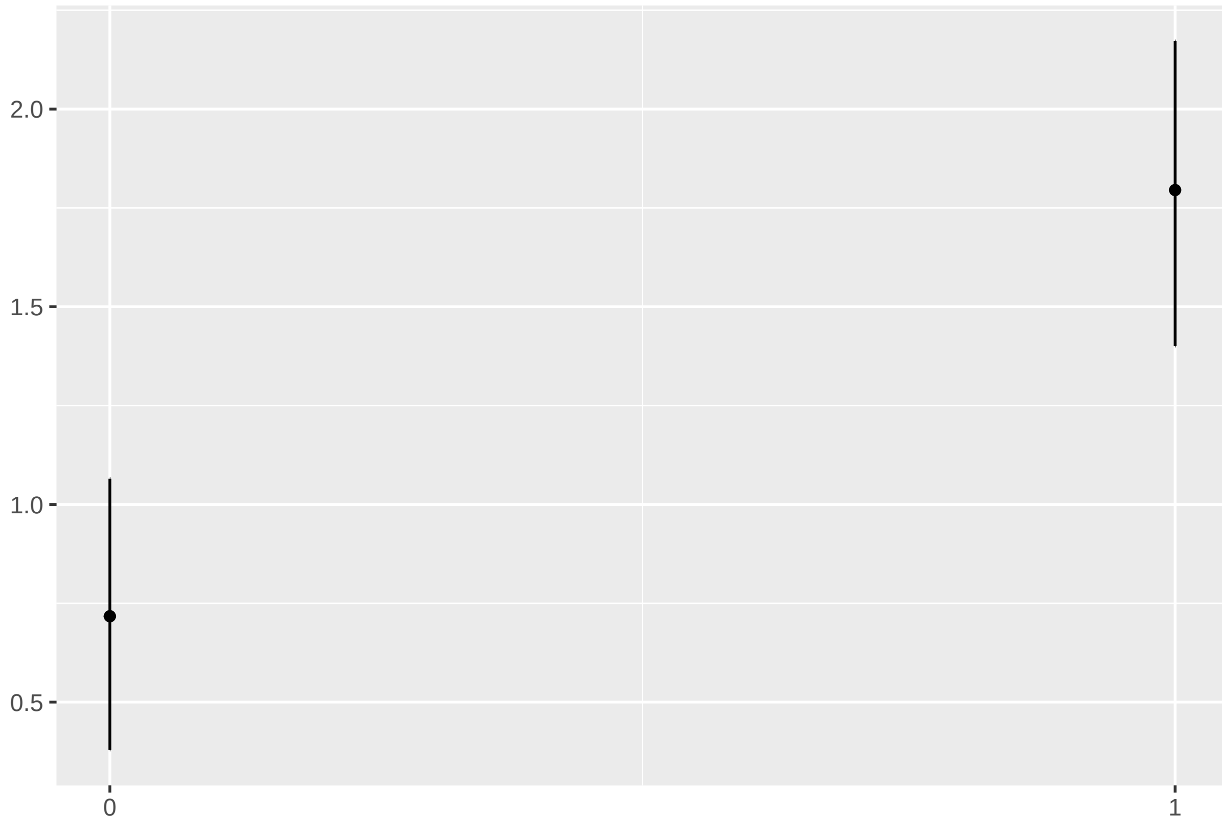
2024-11-04

Interpreting Coefficients

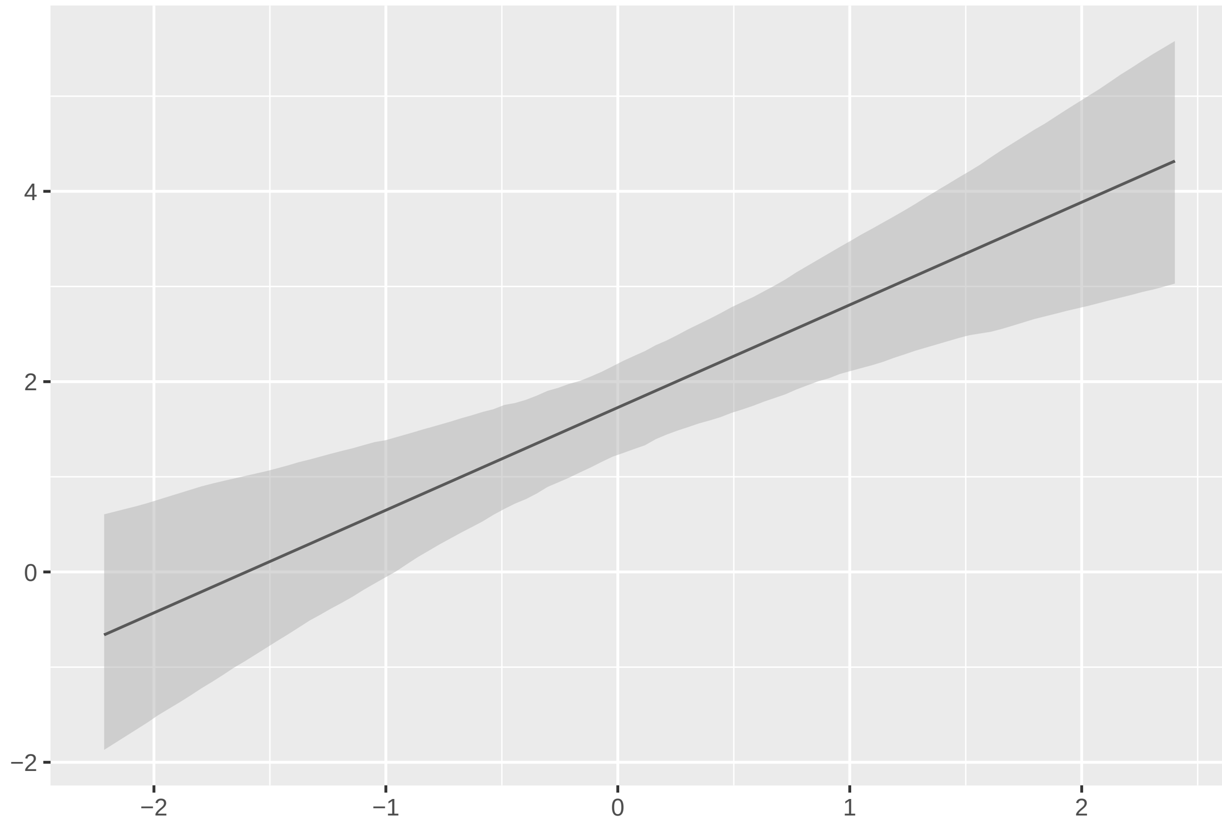
```r
# continuous treatment (X), binary moderator (Z) ##############################
set.seed(1)

n=100

b0=0
b1=1
b2=2
b3=1

data<-data.frame(X=rnorm(n),
                 #rbinom(observations, number of Bernoulli variables, probability)
                 Z=rbinom(n,1,0.5)) #Z = {0,1}

#Y = 1 + 1*X + 2*Z + 1*X*Z + e
for(i in 1:n){
  data$Y[i]<-b0+b1*data$X[i]+b2*data$Z[i]+b3*data$X[i]*data$Z[i]+rnorm(n)[i]
}

par(mfrow=c(1,2))
plot(data$X, data$Y)
plot(data$Z, data$Y)

model<-lm(Y~X+Z+X*Z, data)

library(interplot)
```



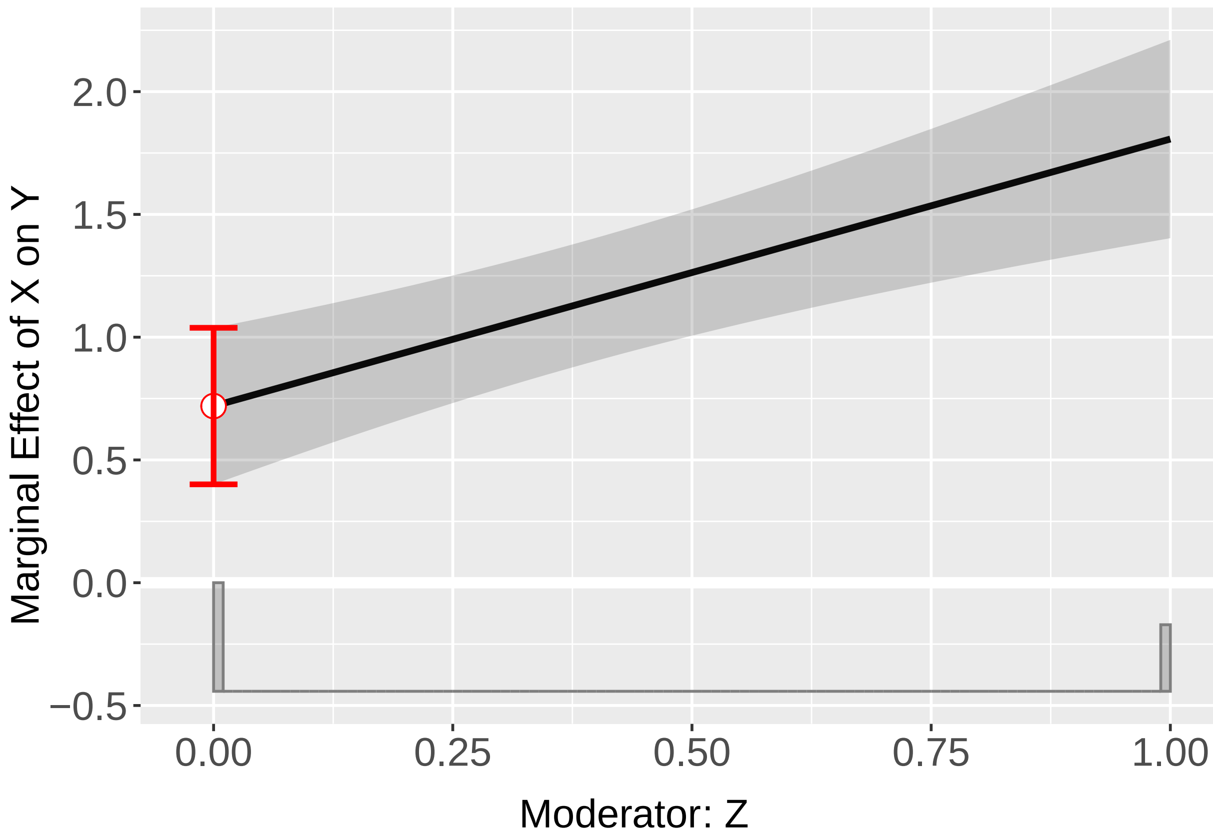

```r
#interplot(model, "variable being moderated", "variable doing the moderating")
interplot(model,"X","Z")
```



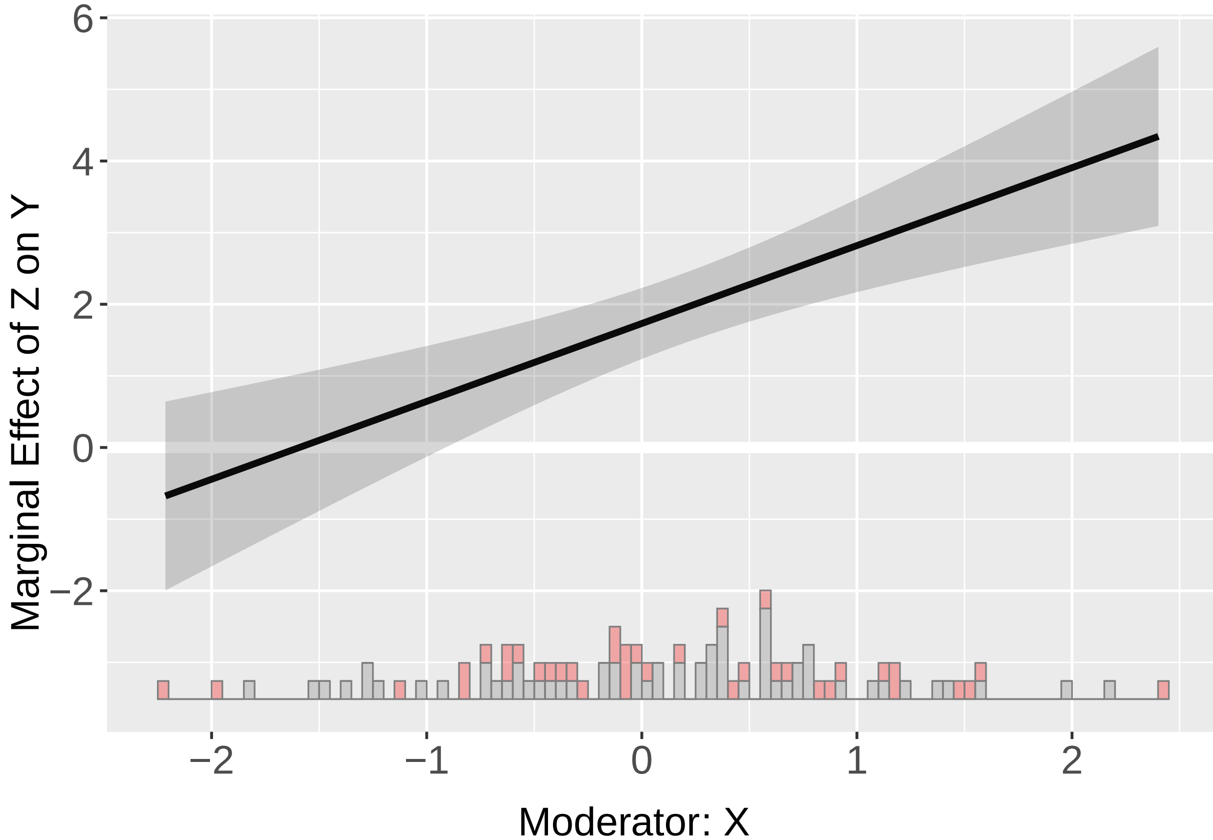

```r
interplot(model,"Z","X")
```



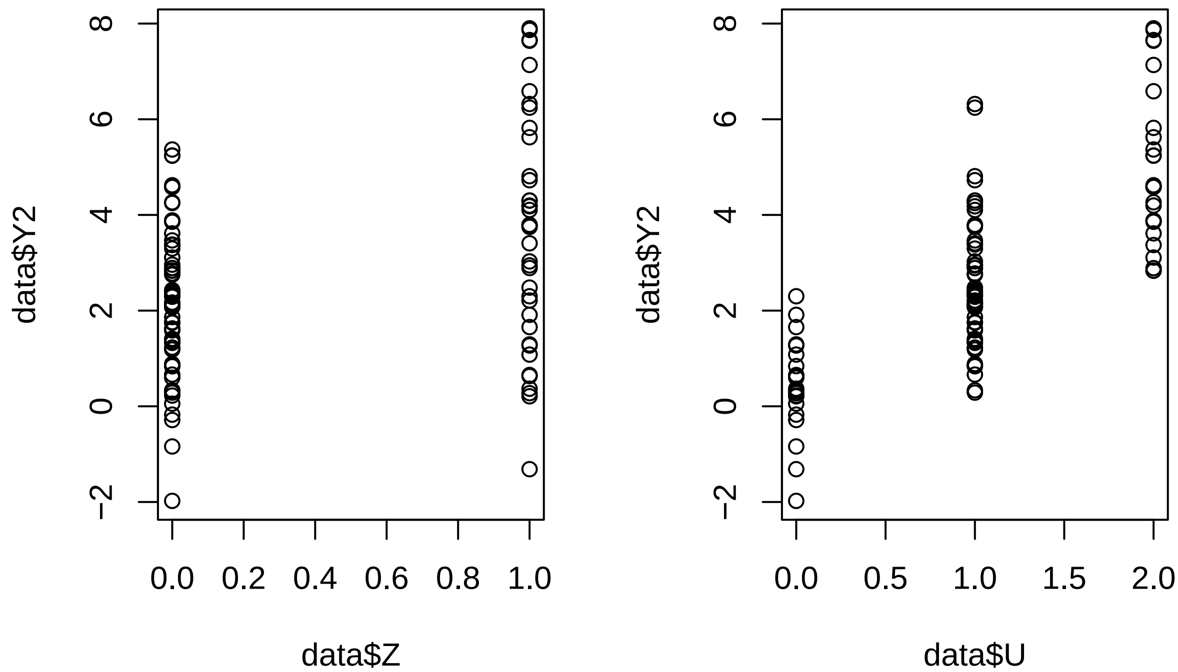

```r
library(interflex)
#interflex("estimator", data, "outcome", "treatment", "moderator")
interflex(estimator="binning",data,"Y","X","Z")
```



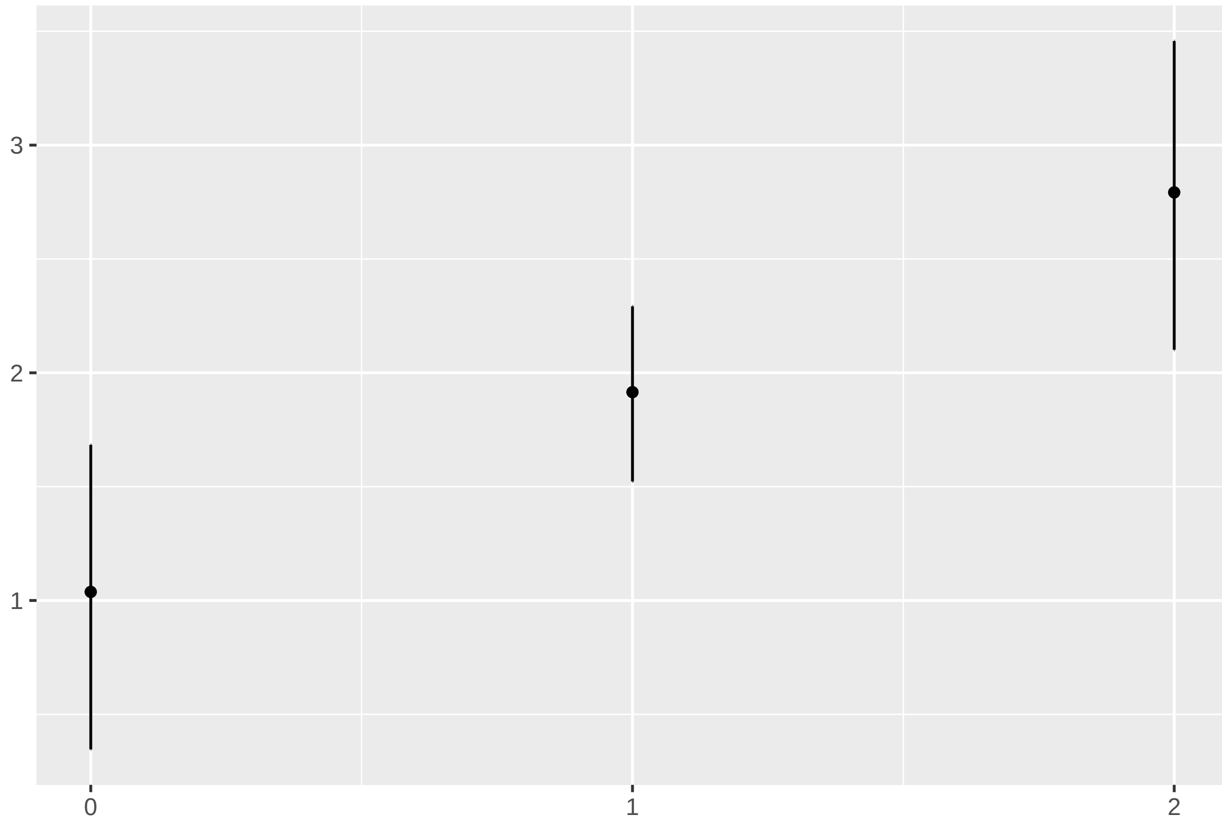

```
interflex(estimator="linear",data,"Y","Z","X")
```



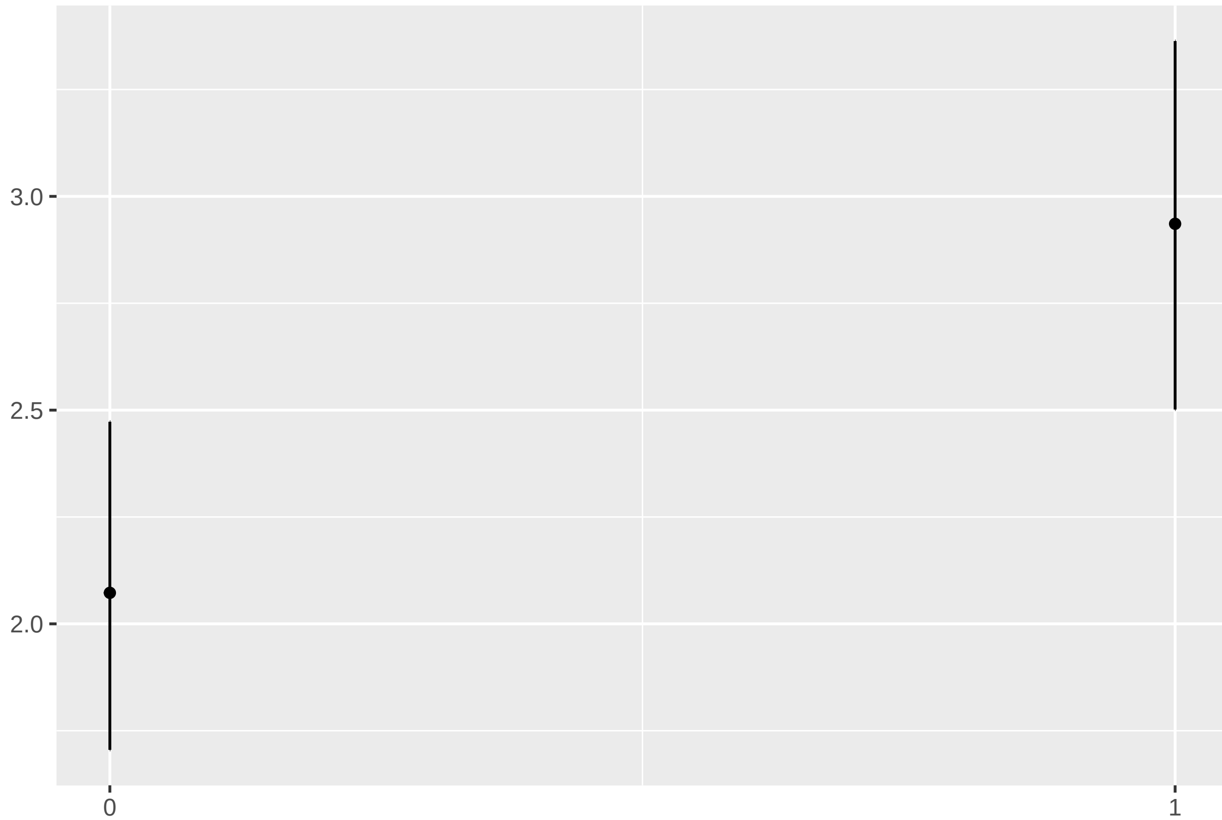

```r
# binary treatment (Z), categorical moderator (U) ##############################
data$U<-rbinom(n,2,0.5) #U = {poor=0, middle class=1, rich=2}

#Y = 1 + 1*Z + 2*U + 1*Z*U + e
for(i in 1:n){
  data$Y2[i]<-b0+b1*data$Z[i]+b2*data$U[i]+b3*data$Z[i]*data$U[i]+rnorm(n)[i]
}

par(mfrow=c(1,2))
plot(data$Z, data$Y2)
plot(data$U, data$Y2)
```



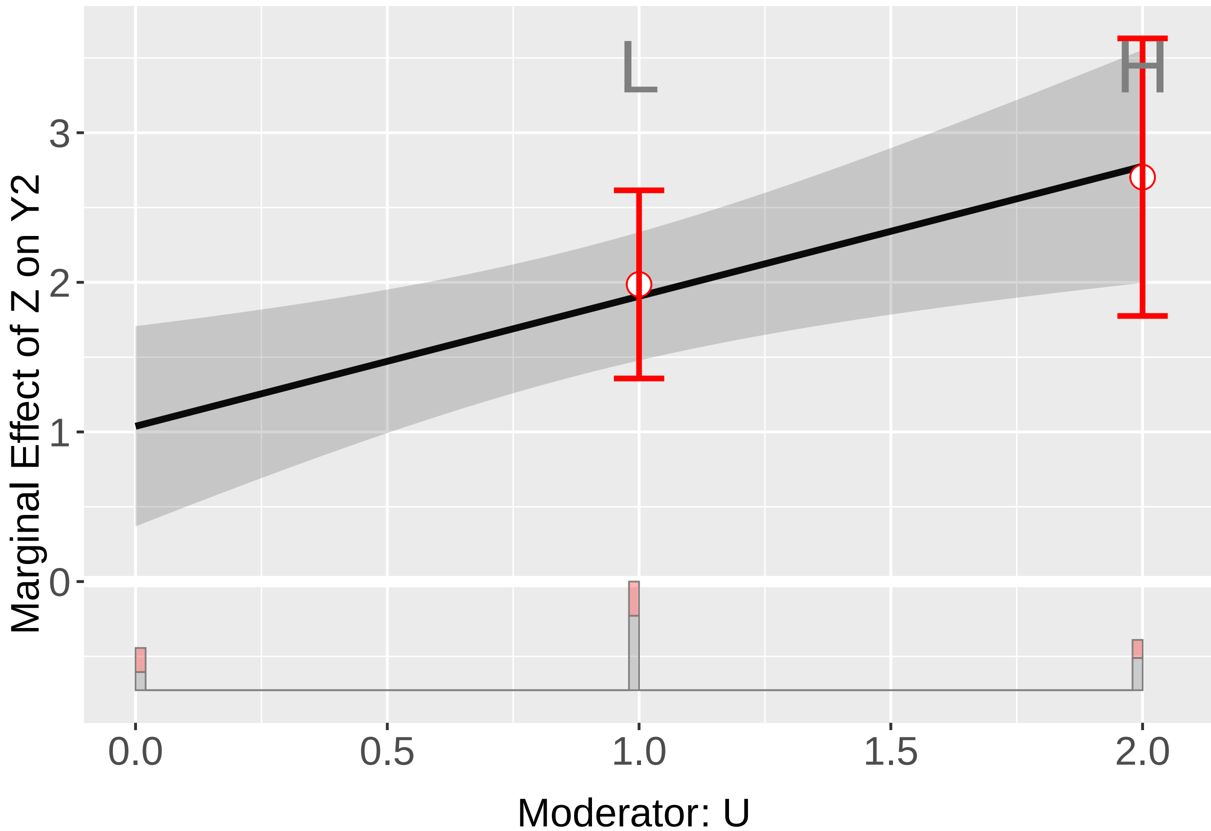

```r
model2<-lm(Y2~Z+U+Z*U, data)

interplot(model2,"Z","U")
```




```
interplot(model2,"U","Z")
```




```r
interflex("binning",data,"Y2","Z","U",nbins=3)
```




```
# binary treatment (Z), "binned" covariates (G, K) ############################
data$G<-ifelse(data$X>0.5,1,0)
data$K<-ifelse(data$X<=0.5,1,0)

#Y = 1 + 1*Z + 2*U + 1*Z*U + e
for(i in 1:n){
  data$Y3[i]<-1+1*data$Z[i]+1*data$G[i]+1*data$K[i]+
    1*data$Z[i]*data$G[i]+rnorm(n)[i]
}

X<-matrix(c(rep(1,n),
            data$Z,
            data$G,
            data$K),ncol=4)
qr(X) #the intercept, G, and K are perfectly collinear

# option 1: omit one category (here `R` dropped K)
model3<-lm(Y3~Z+G+K+Z*G,data)

# option 2: drop intercept
model4<-lm(Y3~-1+Z+G+K+Z*G,data)
```

Sidenote on the `here` package

```r
## quick review of here() ########################################################
library(here)

write.csv(data,file=here("data","data.csv"),row.names=F)

read.csv(here("data","data.csv"))
```

Asymptotic Normality of $\hat{\beta}$

Let's start with arbitrarily looking more closely at $\hat{\beta} - \beta$:

$$
\begin{aligned}
\hat{\beta} - \beta &= (X'X)^{-1}X'Y - \beta \\
&= (X'X)^{-1}X'(X\beta + e) - \beta \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'e - \beta \\
&= (X'X)^{-1}X'e \\
&= (\frac{1}{n}\sum_{i=1}^{n} X_iX_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_ie_i) \\
&= \hat{Q}_{XX}^{-1}\hat{Q}_{Xe} \xrightarrow{p} 0 \implies \hat{\beta} \xrightarrow{p} \beta
\end{aligned}
$$

We can see that $\hat{\beta} - \beta$ can be written as $(\frac{1}{n}\sum_{i=1}^{n} X_iX_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_ie_i)$, which can again be written as $\hat{Q}_{XX}^{-1}\hat{Q}_{Xe}$. Because $\hat{Q}_{Xe} \xrightarrow{p} \mathbb{E}[Xe] = 0$ due to the weak law of large numbers (WLLN), $\hat{\beta} - \beta$ converges in probability to a degenerate distribution on 0. It could also be noted that this implies that $\hat{\beta}$ converges in probability to β, or is consistent, since we subtracted it out $(0 + \beta = \beta)$.

However, we can normalize $\hat{\beta} - \beta$ such that it is not asymptotically degenerate. This is what we did in our `09_16_MSE_WLLN_CLT` lab when we compared non-normalized asymptotic sampling distributions to normalized ones.

When we normalize $\hat{\beta} - \beta$, it no longer converges to a spike about 0 but to a normal distribution centered on 0. Again, this happens due to the relationship between X_i and e_i:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n}\big((X'X)^{-1}X'e\big) && (1) \\
&= \big(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\big)^{-1} \underbrace{\big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_ie_i\big)} && (2) \\
&\qquad\qquad\qquad\qquad \xrightarrow{d} N(0, \Omega) && (3) \\
&\xrightarrow{d} \big(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\big)^{-1} N(0, \Omega) && (4) \\
&\xrightarrow{d} Q_{XX}^{-1} N(0, \Omega) && (5) \\
&\xrightarrow{d} N(0, Q_{XX}^{-1}\Omega Q_{XX}^{-1}) && (6)
\end{aligned}
$$

Here, multiplying $\hat{\beta} - \beta$ by $\sqrt{n}$ normalizes it. Distributing $\sqrt{n}$, we end up with $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e_i$ in the second part of line 2, which line 3 states converges in distribution to a normal with mean 0 and variance Ω. This is because Xe is mean zero ($\mathbb{E}[Xe] = 0$) with variance $\mathbb{E}[(Xe)(Xe)'] = \Omega$. The version of the central limit theorem (CLT) invoked is the (multivariate) Lindeberg-Levy CLT, which states that a sum of mean zero, normalized random vectors converges to a normal: $\sqrt{n}(\overline{Y} - \mu) \xrightarrow{d} N(0, V)$ (Hansen 2022, 160). Here, the "sum of mean zero, normalized random vectors" is $\frac{1}{\sqrt{n}} Xe$ and our V is Ω. The continuous mapping theorem (CMT) was used going from lines 4 to 5 to say $\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i'\right)^{-1} \xrightarrow{p} Q_{XX}^{-1}$ and Slutsky's theorem was used going from lines 5 to 6.

Thus, the asymptotic variance of is $\mathbb{V}[\beta] = Q_{XX}^{-1} \Omega Q_{XX}^{-1}$. This might seem (slightly?) more intuitive if you consider what we learned last week about the finite variance of $\hat{\beta}$:

$$\mathbb{V}[\hat{\beta}|X] = \mathbb{E}\Big[\big((X'X)^{-1} \underbrace{X'e\big)\big(e'X}_{\text{looks a lot like } \mathbb{E}[(Xe)(Xe)'] = \Omega}(X'X)^{-1}\big)|X\Big]$$

Indeed, if we multiply $\mathbb{V}[\hat{\beta}|X]$ by n, we get a consistent expression of the asymptotic variance:

$$\begin{aligned}
n\mathbb{V}[\hat{\beta}|X] &= n\left(\mathbb{E}\Big[\big((X'X)^{-1}X'e\big)\big(e'X(X'X)^{-1}\big)|X\Big]\right) \\
&= \big(\frac{1}{n}(X'X)^{-1}\big)\ \big(\frac{1}{n}X'(\mathbb{E}[ee'|X])X\big)\ \big(\frac{1}{n}(X'X)^{-1}\big) \\
&= \hat{Q}_{XX}^{-1}\ \big(\frac{1}{n}X'DX\big)\ \hat{Q}_{XX}^{-1} \\
&\xrightarrow{p} Q_{XX}^{-1}\Omega Q_{XX}^{-1}
\end{aligned}$$

In conclusion, we can approximate the sampling distribution of our OLS estimator as: $\hat{\beta}|X \sim N(\beta, \mathbb{V}[\beta]/n)$ (Blackwell).

Bibliography

Hansen, Bruce. 2022. *Econometrics.* Princeton University Press.