Marginal and conditional effects

Peter Hoff

STAT 423

Applied Regression and Analysis of Variance

University of Washington

Possibilities

If the marginal effect is $+$,significant, the conditional effect may be

- $+$ and significant
- $+$ and nonsignificant
- - and significant
- - and nonsignificant

Possibilities

If the marginal effect is $+$,significant, the conditional effect may be

- $+$ and significant
- $+$ and nonsignificant
- - and significant
- - and nonsignificant

If the marginal effect is $+$,nonsignificant, the conditional effect may be

- $+$ and significant
- $+$ and nonsignificant
- - and significant
- - and nonsignificant

Possibilities

If the marginal effect is $+$,significant, the conditional effect may be

- $+$ and significant
- $+$ and nonsignificant
- - and significant
- - and nonsignificant

If the marginal effect is $+$,nonsignificant, the conditional effect may be

- $+$ and significant
- $+$ and nonsignificant
- - and significant
- - and nonsignificant

In other words, neither the sign nor significance of the effects need match.

Marginal and conditional effects

Marginal effect of x_1:

- Model: $y = \beta_{01} + \beta_1 x_1 + \epsilon$
- Effect: $\hat{\beta}_1 = SX_1Y/SX_1X_1$

Marginal and conditional effects

Marginal effect of x_1:

- Model: $y = \beta_{01} + \beta_1 x_1 + \epsilon$
- Effect: $\hat{\beta}_1 = SX_1Y/SX_1X_1$

Marginal effect of x_2:

- Model: $y = \beta_{02} + \beta_2 x_2 + \epsilon$
- Effect: $\hat{\beta}_2 = SX_2Y/SX_2X_2$

Marginal and conditional effects

Marginal effect of x_1:

- Model: $y = \beta_{01} + \beta_1 x_1 + \epsilon$
- Effect: $\hat{\beta}_1 = SX_1Y/SX_1X_1$

Marginal effect of x_2:

- Model: $y = \beta_{02} + \beta_2 x_2 + \epsilon$
- Effect: $\hat{\beta}_2 = SX_2Y/SX_2X_2$

Conditional effects:

- Model: $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \epsilon$
- Effect: $\hat{\alpha}_1 = \left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right]_{[2]}$

How are $\hat{\beta}_1$ and $\hat{\alpha}_1$ related?

Last time, we showed that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are centered, then

$$\hat{\beta}_1 = \mathbf{x}_1^T \mathbf{y} / \mathbf{x}_1^T \mathbf{x}_1$$

Last time, we showed that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are centered, then

$$\hat{\beta}_1 = \mathbf{x}_1^T \mathbf{y} / \mathbf{x}_1^T \mathbf{x}_1$$

$$\hat{\alpha}_1 = \left((\mathbf{x}_2^T \mathbf{x}_2) \mathbf{x}_1^T \mathbf{y} - (\mathbf{x}_1^T \mathbf{x}_2) \mathbf{x}_2^T \mathbf{y} \right) / w \ , \ \text{where}$$
$$w = \mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2.$$

We can rewrite this as

$$\begin{aligned}
\hat{\alpha}_1 &= \frac{\mathbf{x}_1^T \mathbf{y}}{\mathbf{x}_1^T \mathbf{x}_1} \times \frac{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} - \frac{\mathbf{x}_2^T \mathbf{y}}{\mathbf{x}_2^T \mathbf{x}_2} \times \frac{\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}_1^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} \\
&= \hat{\beta}_1 \times \frac{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} - \hat{\beta}_2 \times \frac{\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}_1^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2}
\end{aligned}$$

$$\hat{\alpha}_1 = \quad \hat{\beta}_1 \quad \times \frac{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} \; - \quad \hat{\beta}_2 \quad \times \frac{\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}_1^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2}$$

Q: How does $\mathbf{x}_1^T \mathbf{x}_2$ affect the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1$?

$$\hat{\alpha}_1 = \quad \hat{\beta}_1 \quad \times \frac{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2}{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2 - (\mathbf{x}_1^T\mathbf{x}_2)^2} \; - \quad \hat{\beta}_2 \quad \times \frac{\mathbf{x}_2^T\mathbf{x}_2\mathbf{x}_1^T\mathbf{x}_2}{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2 - (\mathbf{x}_1^T\mathbf{x}_2)^2}$$

Q: How does $\mathbf{x}_1^T\mathbf{x}_2$ affect the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1$?

Result 1: If $\mathbf{x}_1^T\mathbf{x}_2 = 0$, $\hat{\alpha}_1 = \hat{\beta}_1$.

$$\hat{\alpha}_1 = \quad \hat{\beta}_1 \quad \times \frac{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} \; - \quad \hat{\beta}_2 \quad \times \frac{\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}_1^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2}$$

Q: How does $\mathbf{x}_1^T \mathbf{x}_2$ affect the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1$?

Result 1: If $\mathbf{x}_1^T \mathbf{x}_2 = 0$, $\hat{\alpha}_1 = \hat{\beta}_1$.

Q: What if $\mathbf{x}_1^T \mathbf{x}_2 \neq 0$?

$$\hat{\alpha}_1 = \quad \hat{\beta}_1 \quad \times \frac{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2} \; - \quad \hat{\beta}_2 \quad \times \frac{\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}_1^T \mathbf{x}_2}{\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{x}_2 - (\mathbf{x}_1^T \mathbf{x}_2)^2}$$

Q: How does $\mathbf{x}_1^T \mathbf{x}_2$ affect the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1$?

Result 1: If $\mathbf{x}_1^T \mathbf{x}_2 = 0$, $\hat{\alpha}_1 = \hat{\beta}_1$.

Q: What if $\mathbf{x}_1^T \mathbf{x}_2 \neq 0$?

Result 2: Let $\mathbf{x}_1^T \mathbf{x}_1 = \mathbf{x}_2^T \mathbf{x}_2 = 1$, and let $\rho = \mathbf{x}_1^T \mathbf{x}_2 = \text{Cor}(\mathbf{x}_1, \mathbf{x}_2)$.

$$\hat{\alpha}_1 = \quad \hat{\beta}_1 \frac{1}{1 - \rho^2} - \hat{\beta}_2 \frac{\rho}{1 - \rho^2}$$

For ρ close to 1, this is a large linear combination of $\hat{\beta}_1, \hat{\beta}_2$.

$$\hat{\alpha}_1 = \hat{\beta}_1 \times \frac{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2}{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2 - (\mathbf{x}_1^T\mathbf{x}_2)^2} - \hat{\beta}_2 \times \frac{\mathbf{x}_2^T\mathbf{x}_2\mathbf{x}_1^T\mathbf{x}_2}{\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_2 - (\mathbf{x}_1^T\mathbf{x}_2)^2}$$

Q: How does $\mathbf{x}_1^T\mathbf{x}_2$ affect the relationship between $\hat{\alpha}_1$ and $\hat{\beta}_1$?

Result 1: If $\mathbf{x}_1^T\mathbf{x}_2 = 0$, $\hat{\alpha}_1 = \hat{\beta}_1$.

Q: What if $\mathbf{x}_1^T\mathbf{x}_2 \neq 0$?

Result 2: Let $\mathbf{x}_1^T\mathbf{x}_1 = \mathbf{x}_2^T\mathbf{x}_2 = 1$, and let $\rho = \mathbf{x}_1^T\mathbf{x}_2 = \mathrm{Cor}(\mathbf{x}_1, \mathbf{x}_2)$.

$$\hat{\alpha}_1 = \hat{\beta}_1 \frac{1}{1-\rho^2} - \hat{\beta}_2 \frac{\rho}{1-\rho^2}$$

For ρ close to 1, this is a large linear combination of $\hat{\beta}_1, \hat{\beta}_2$.

What do you think this does to the variance of $\hat{\alpha}_1$?

Numerical example

True model:

$$y = .25x_1 + .75x_2 + \epsilon$$

Numerical example

True model:

$$y = .25x_1 + .75x_2 + \epsilon$$

Collinearity:

$$x_2 = \rho \times x_1 + \sqrt{(}1 - \rho^2) \times z$$

Then if $\mathrm{Var}[x_1] = \mathrm{Var}[z]$, $\mathrm{Cor}[x_1, x_2] = \rho$.

$\rho = 0$

```
rho<-.0

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.02628723
```

$\rho = 0$

```
rho<-.0

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.02628723
```

```
y<- .25*x1 + .75*x2 + e

lm(y~x1)$coef

## (Intercept)          x1
## 0.005394464 0.225787641

lm(y~x2)$coef

## (Intercept)          x2
##   0.1439477   0.7598974

summary(lm(y~x1+x2))$coef

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.1226683  0.1429228 0.8582832 3.950921e-01
## x1          0.2043111  0.1699043 1.2025070 2.351904e-01
## x2          0.7549346  0.1569981 4.8085576 1.602073e-05
```

$\rho = .5$

```
rho<-.5

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.4885865
```

$\rho = .5$

```
rho<-.5

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.4885865
```

```
y<- .25*x1 + .75*x2 + e

lm(y~x1)$coef

## (Intercept)          x1
##  0.02100348  0.59792914

lm(y~x2)$coef

## (Intercept)          x2
##   0.1504062   0.8473659

summary(lm(y~x1+x2))$coef

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.1226683  0.1429228 0.8582832 0.3950921151
## x1          0.2014621  0.1946620 1.0349332 0.3059959623
## x2          0.7556979  0.1812858 4.1685440 0.0001305312
```

$\rho = .95$

```
rho<-.95

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.9431838
```

$\rho = .95$

```
rho<-.95

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 0.9431838
```

```
y<- .25*x1 + .75*x2 + e

lm(y~x1)$coef

## (Intercept)          x1
##  0.08552234  0.92361371

lm(y~x2)$coef

## (Intercept)          x2
##   0.1332862   0.9414229

summary(lm(y~x1+x2))$coef

##              Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) 0.1226683  0.1429228 0.8582832 0.3950921
## x1          0.1892981  0.5111654 0.3703264 0.7128030
## x2          0.7658032  0.5027964 1.5230881 0.1344374
```

$\rho = 1$

```
rho<-1

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 1
```

$\rho = 1$

```r
rho<-1

x2<-rho*x1 + sqrt(1-rho^2)*z

cor(x1,x2)

## [1] 1
```

```r
y<- .25*x1 + .75*x2 + e

lm(y~x1)$coef

## (Intercept)          x1
##   0.1219017   0.9544515

lm(y~x2)$coef

## (Intercept)          x2
##   0.1219017   0.9544515

summary(lm(y~x1+x2))$coef

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.1219017  0.1393534 0.8747666 3.860570e-01
## x1          0.9544515  0.1680688 5.6789325 7.726316e-07
```

Results

As correlation increased,

- estimated conditional effects stayed about the same;
- standard errors increased, t-values decreased, p-values increased;
- estimated marginal effects increased.

Results

As correlation increased,

- estimated conditional effects stayed about the same;
- standard errors increased, t-values decreased, p-values increased;
- estimated marginal effects increased.

This latter phenomonenon is called **omitted variable bias**.

Results

As correlation increased,

- estimated conditional effects stayed about the same;
- standard errors increased, t-values decreased, p-values increased;
- estimated marginal effects increased.

This latter phenomonenon is called **omitted variable bias**.

At $\rho = 1$,

- the design matrix $\mathbf{X}$ is **rank deficient**;
- the model is **overparameterized**.

R and many other packages simply drop one of the redundant variables.

Omitted variable bias

True model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\gamma} + \mathbf{e}$

Omitted variable bias

True model: $\mathbf{y} = \mathbf{X}_1\beta + \mathbf{X}_2\gamma + \mathbf{e}$

Fitted model: $\mathbf{y} = \mathbf{X}_1\beta + \mathbf{e}$

Omitted variable bias

True model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\gamma + \mathbf{e}$

Fitted model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{e}$

Omitted variable bias:

$$\begin{aligned} \mathrm{E}[\hat{\boldsymbol{\beta}}_1|\mathbf{X}_1, \mathbf{X}_2] &= (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T \mathrm{E}[\mathbf{y}|\mathbf{X}_1, \mathbf{X}_2] \\ &= (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T(\mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\gamma) \\ &= \boldsymbol{\beta}_1 + (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{X}_2\gamma \end{aligned}$$

If $\mathbf{X}_2$ is "fixed":
The bias is zero only if $\mathbf{X}_1^T\mathbf{X}_2 = 0$.

Omitted variable bias

True model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\gamma + \mathbf{e}$

Fitted model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{e}$

Omitted variable bias:

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\beta}}_1|\mathbf{X}_1, \mathbf{X}_2] &= (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathrm{E}[\mathbf{y}|\mathbf{X}_1, \mathbf{X}_2] \\
&= (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T(\mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\gamma) \\
&= \boldsymbol{\beta}_1 + (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{X}_2\gamma
\end{aligned}$$

If $\mathbf{X}_2$ is "fixed":
The bias is zero only if $\mathbf{X}_1^T\mathbf{X}_2 = 0$.

If $\mathbf{X}_2$ is "random":
The bias (on average across $\mathbf{X}_2$) is zero only if $\mathrm{E}[\mathbf{X}_1^T\mathbf{X}_2|\mathbf{X}_1] = 0$.

Marginal versus conditional effects

x_1, x_2 correlated $\Rightarrow$ marginal, conditional effects are generally different.

Marginal versus conditional effects

x_1, x_2 correlated $\Rightarrow$ marginal, conditional effects are generally different.

Q: Which effects are of most scientific interest?

Marginal versus conditional effects

x_1, x_2 correlated $\Rightarrow$ marginal, conditional effects are generally different.

Q: Which effects are of most scientific interest?

A: It depends on the situation.

Lurking variable

Consider a study of a sample of flu patients:

- Some patients choose to buy an expensive new drug, others don't.
- Illness severity is recorded for all patients in the study.

Let

- x_1 indicate whether or not a patient takes the drug;
- y denote disease severity.

Suppose a linear regression shows a significant decrease in severity among those taking the drug.

Lurking variable

Consider a study of a sample of flu patients:

- Some patients choose to buy an expensive new drug, others don't.
- Illness severity is recorded for all patients in the study.

Let

- x_1 indicate whether or not a patient takes the drug;
- y denote disease severity.

Suppose a linear regression shows a significant decrease in severity among those taking the drug.

Would you take the drug if you were sick?

Lurking variable

Suppose

- there were no effect of taking the drug on disease;
- wealthier people were more likely to take the drug;
- wealthier people had better overall healthcare.

Lurking variable

Suppose

- there were no effect of taking the drug on disease;
- wealthier people were more likely to take the drug;
- wealthier people had better overall healthcare.

In other words,

- the *conditional effect* of x_1 on y given x_2 is zero;
- x_1 and x_2 are correlated;
- the *conditional effect* of x_2 on y given x_1 is nonzero.
- the *marginal effect* of x_1 on y is nonzero.

Lurking variable

Suppose

- there were no effect of taking the drug on disease;
- wealthier people were more likely to take the drug;
- wealthier people had better overall healthcare.

In other words,

- the *conditional effect* of x_1 on y given x_2 is zero;
- x_1 and x_2 are correlated;
- the *conditional effect* of x_2 on y given x_1 is nonzero.
- the *marginal effect* of x_1 on y is nonzero.

Are marginal or conditional effects of the drug more relevant?

Lurking variable

Suppose

- there were no effect of taking the drug on disease;
- wealthier people were more likely to take the drug;
- wealthier people had better overall healthcare.

In other words,

- the *conditional effect* of x_1 on y given x_2 is zero;
- x_1 and x_2 are correlated;
- the *conditional effect* of x_2 on y given x_1 is nonzero.
- the *marginal effect* of x_1 on y is nonzero.

Are marginal or conditional effects of the drug more relevant?

Which would help you decide whether or not to take the drug?

Indirect effects

Consider a study of an exercise program on blood pressure.

- Subjects randomly assigned to an exercise program or control.
- Blood pressure and other variables measured at the end of the study.

Indirect effects

Consider a study of an exercise program on blood pressure.

- Subjects randomly assigned to an exercise program or control.
- Blood pressure and other variables measured at the end of the study.

Let

- x_1 indicate being on the exercise program or not;
- x_2 denote change in BMI;
- y denote change in blood pressure.

Indirect effects

$\hat{\alpha}_1$: OLS estimate from $y \sim x_1 + x_2$

$\hat{\beta}_1$: OLS estimate from $y \sim x_1$

Indirect effects

$\hat{\alpha}_1$: OLS estimate from $y \sim x_1 + x_2$

$\hat{\beta}_1$: OLS estimate from $y \sim x_1$

Suppose $\hat{\alpha}_1$ is smaller, less significant than $\hat{\beta}_1$.

Indirect effects

$\hat{\alpha}_1$: OLS estimate from $y \sim x_1 + x_2$

$\hat{\beta}_1$: OLS estimate from $y \sim x_1$

Suppose $\hat{\alpha}_1$ is smaller, less significant than $\hat{\beta}_1$.

Why might this be?

Indirect effects

$\hat{\alpha}_1$: OLS estimate from $y \sim x_1 + x_2$

$\hat{\beta}_1$: OLS estimate from $y \sim x_1$

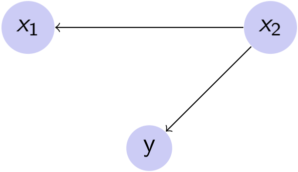
Suppose $\hat{\alpha}_1$ is smaller, less significant than $\hat{\beta}_1$.

Why might this be?

Which effect is more relavent for assessing effect of the exercise regimen?

Graphical models

Drug example:



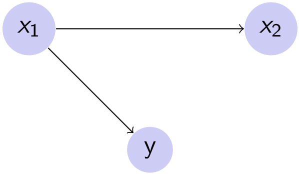
x_1 has no *causal* effect

x_1 has no *conditional* effect

x_1 has a *marginal* effect

Graphical models

Exercise example:



x_1 has an indirect causal effect

x_1 may have a direct causal effect

marginal effect of x_1 $\approx$ direct + indirect causal effect

conditional effect of x_1 $\approx$ direct causal effect

conditional effect $<$ marginal effect $\approx$ total causal effect