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

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 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

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

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

We can rewrite this as

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

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

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

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 $\rho = \mathbf{x}_1^T\mathbf{x}_2 = \mathrm{Cor}(\mathbf{x}_1, \mathbf{x}_2)$.

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

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

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

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 $\text{Var}[x_1] = \text{Var}[z]$, $\text{Cor}[x_1, x_2] = \rho$.

$\rho = 0$

```r
rho<-.0

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

cor(x1,x2)

## [1] 0.02628723
```

```r
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
```

```
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
```

```
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
```

```
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.

This latter phenomenon 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: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\alpha}_1 + \mathbf{X}_2\boldsymbol{\alpha}_2 + \mathbf{e}$

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

Omitted variable bias:

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

The "bias" is the term $(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\alpha}_2$

However, this is just the difference between the

- true conditional effect $\boldsymbol{\alpha}_1$, and the
- expectation of the estimated marginal effect $\hat{\boldsymbol{\beta}}_1$.

Omitted variable bias:

$$\text{bias} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \boldsymbol{\alpha}_2$$

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

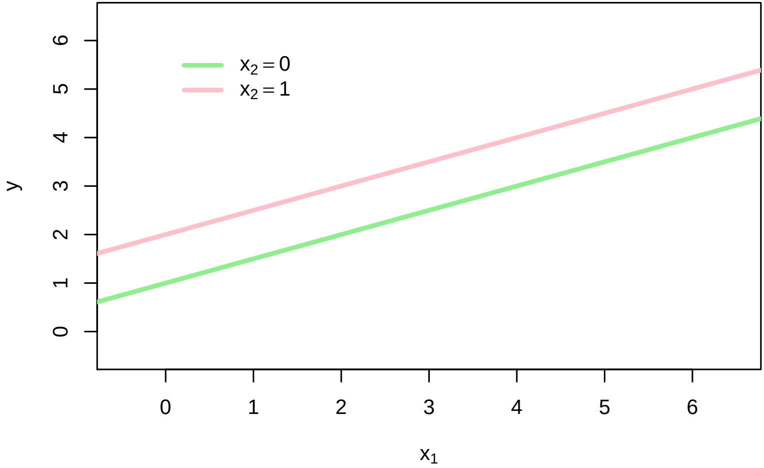
If $\mathbf{X}_2$ is "random":
The bias (on average across $\mathbf{X}_2$) is zero only if $\mathrm{E}[\mathbf{X}_1^T \mathbf{X}_2 | \mathbf{X}_1] = 0$.

This corresponds to what we have discussed before

- $\hat{\alpha}_1 = \hat{\beta}_1$ if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have zero *sample* correlation.
- $\alpha_1 = \beta_1$ if the conditional expectation of x_2 is not linearly dependent on x_1.

Visualizing marginal and conditional effects

$$\mathsf{E}[y|x_1, x_2] = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$



Case 1: x_2 independent of x_1

Suppose $x_2|x_1 \sim \text{binary}(1/3)$

What is $\text{E}[y|x_1]$?

$$\begin{aligned}
\text{E}[y|x_1] &= \text{E}[\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 | x_1] \\
&= \alpha_0 + \alpha_1 x_1 + \alpha_2 \text{E}[x_2|x_1] \\
&= \alpha_0 + \alpha_1 x_1 + \alpha_2/3 \\
&= (\alpha_0 + \alpha_2/3) + \alpha_1 x_1 \\
&\equiv \qquad \beta_0 \qquad\quad + \beta_1 x_1
\end{aligned}$$

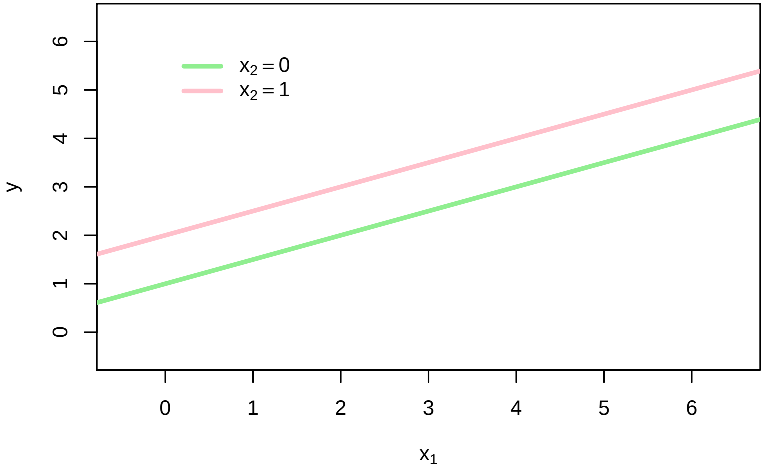
$\text{E}[y|x_1, x_2]$ and $\text{E}[y|x_1]$ are *both* linear in x_1.

(This is because $\text{E}[x_2|x_1]$ is linear in x_1).

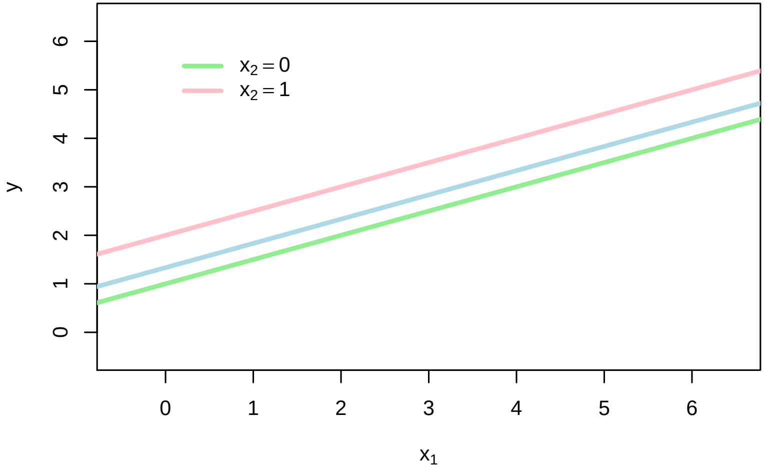
The marginal effect of x_1 is the same its conditional effect.

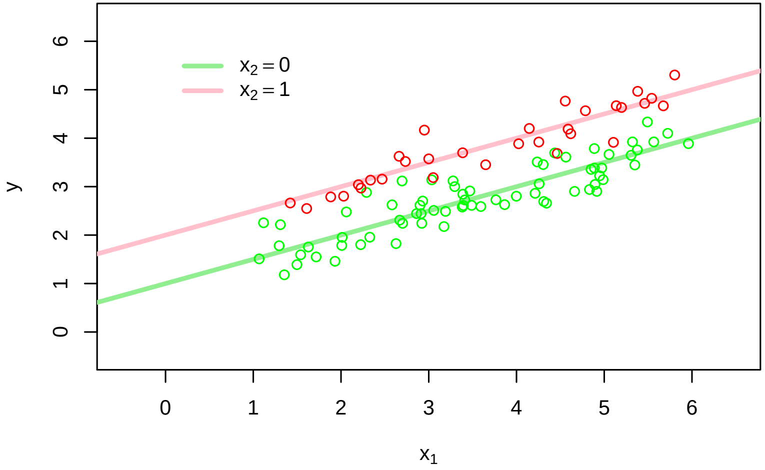
(This is because $\text{E}[x_2|x_1]$ does not depend on x_1).

$$E[y|x_1, x_2] = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$

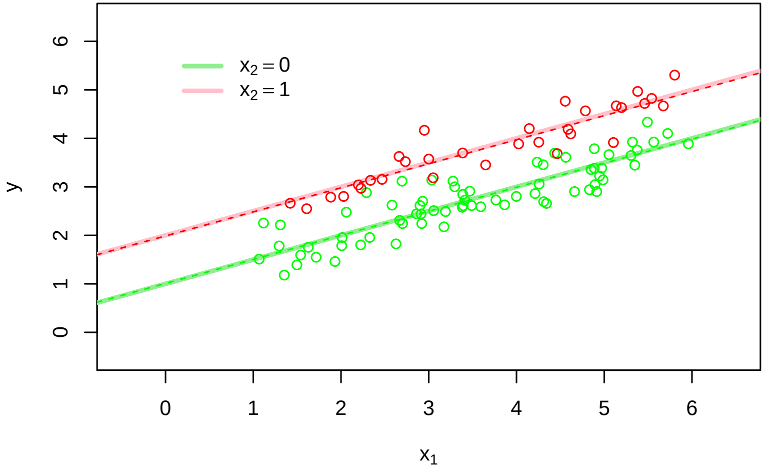


$$E[y|x_1] = (\alpha_0 + \alpha_2/3) + \alpha_1 x_1$$

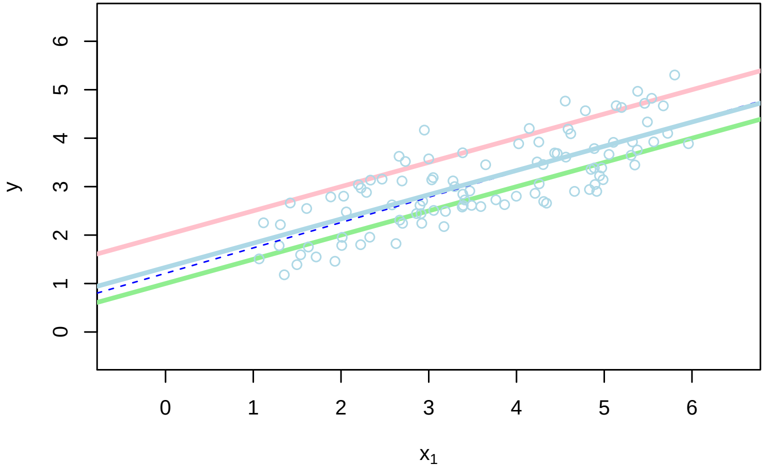




$x_2 = 0$
$x_2 = 1$
y
x_1
0
1
2
3
4
5
6



x2 = 0
x2 = 1
y
x1
0
1
2
3
4
5
6



y
0
1
2
3
4
5
6
0
1
2
3
4
5
6
x_1

Case 2: x_2 linearly dependent on x_1

Suppose $x_2|x_1 \sim \text{binary}(\gamma_0 + \gamma_1 x_1)$

What is $\text{E}[y|x_1]$?

$$\begin{aligned}
\text{E}[y|x_1] &= \text{E}[\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 | x_1] \\
&= \alpha_0 + \alpha_1 x_1 + \alpha_2 \text{E}[x_2|x_1] \\
&= \alpha_0 + \alpha_1 x_1 + \alpha_2 \gamma_0 + \alpha_2 \gamma_1 x_1 \\
&= (\alpha_0 + \alpha_2 \gamma_0) + (\alpha_1 + \alpha_2 \gamma_1) x_1 \\
&\equiv \qquad \beta_0 \qquad\quad + \beta_1 x_1
\end{aligned}$$

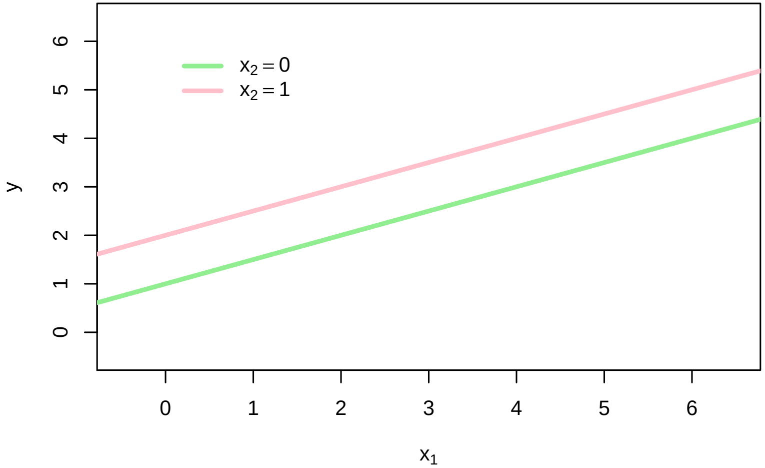
$\text{E}[y|x_1, x_2]$ and $\text{E}[y|x_1]$ are *both* linear in x_1.

(This is because $\text{E}[x_2|x_1]$ is linear in x_1).

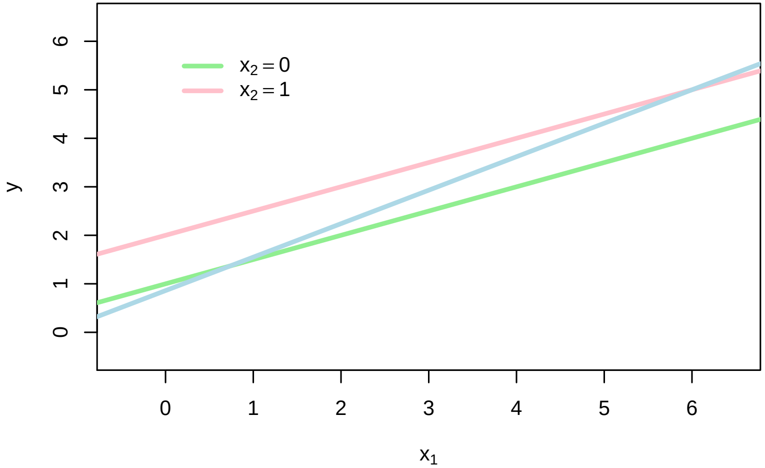
The marginal effect of x_1 is *not* the same its conditional effect.

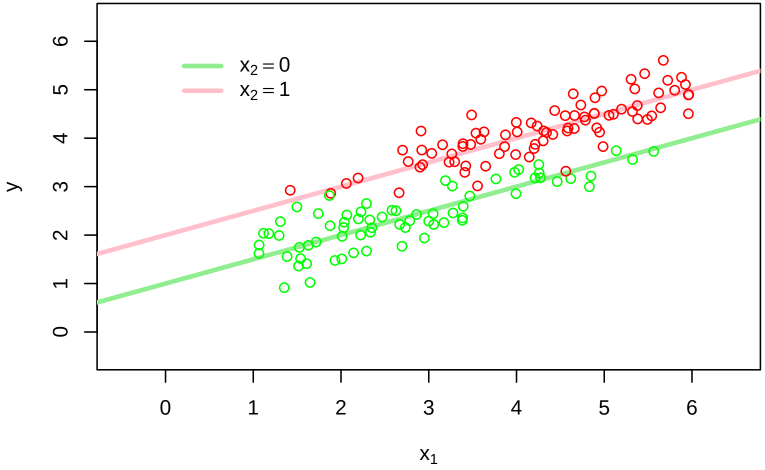
(This is because $\text{E}[x_2|x_1]$ depends on x_1).

$$E[y|x_1, x_2] = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$

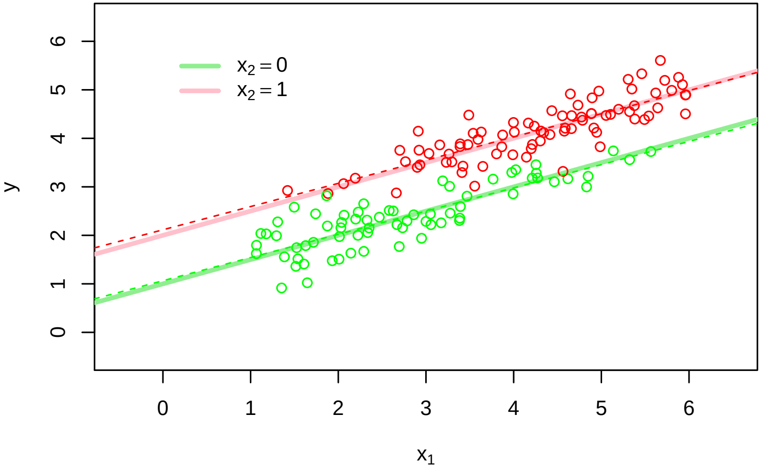


$$E[y|x_1] = (\alpha_0 + \gamma_0) + (\alpha_1 + \gamma_1\alpha_2)x_1$$

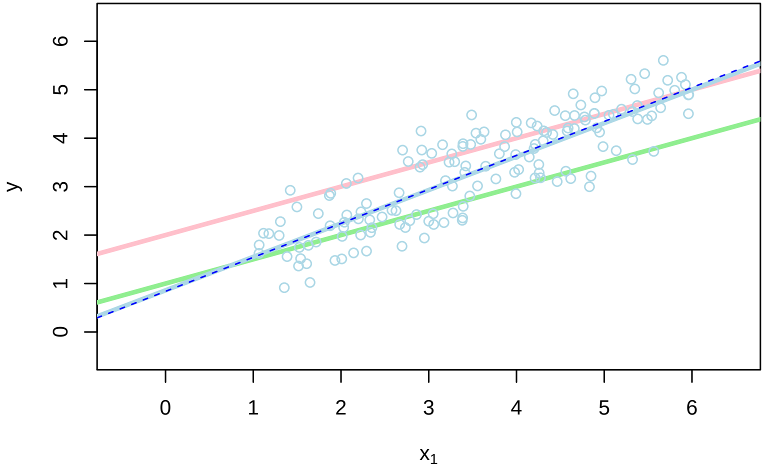




$x_2 = 0$
$x_2 = 1$
y
0
1
2
3
4
5
6
0
1
2
3
4
5
6
x_1



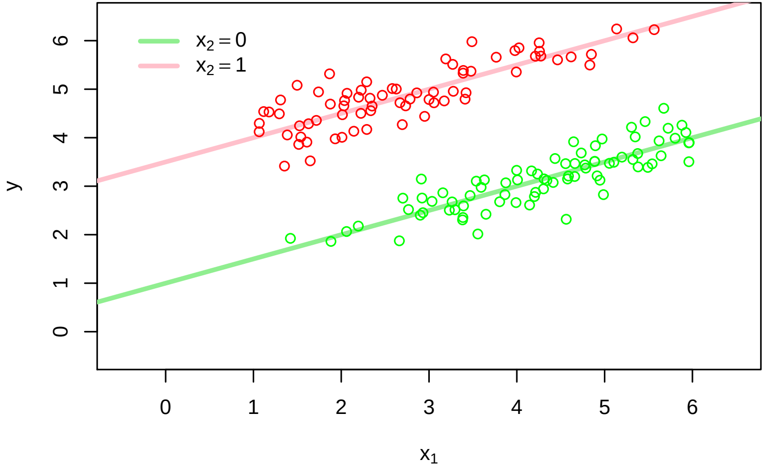
$x_2 = 0$
$x_2 = 1$
y
x_1
0
1
2
3
4
5
6



0
1
2
3
4
5
6
x_1
y

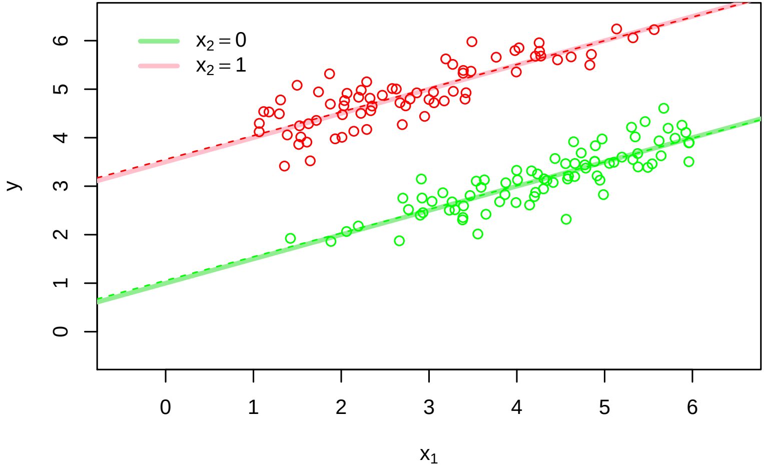
Extreme case:

- large positive effect of x_2
- large negative correlation between x_1, x_2



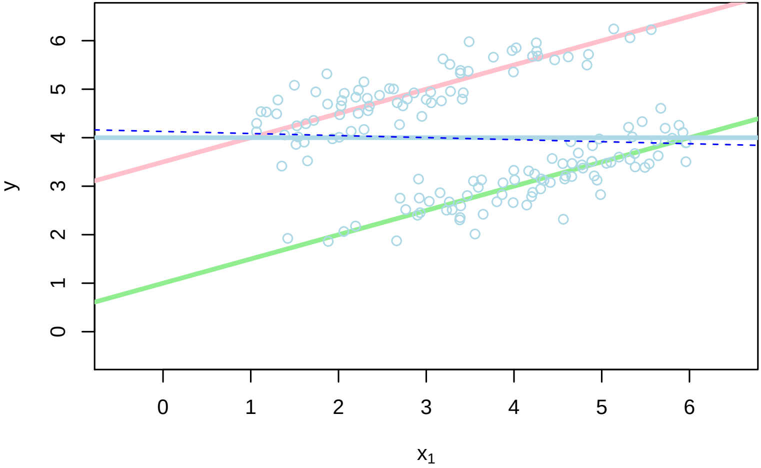
Extreme case:

- large positive effect of x_2
- large negative correlation between x_1, x_2



Extreme case:

- large positive effect of x_2
- large negative correlation between x_1, x_2



Case 3: x_2 nonlinearly dependent on x_1

Suppose $x_2|x_1 \sim \text{binary}(p(x_1) = \frac{\exp(\gamma_0+\gamma_1 x_1)}{1+\exp(\gamma_0+\gamma_1 x_1)})$

What is $E[y|x_1]$?

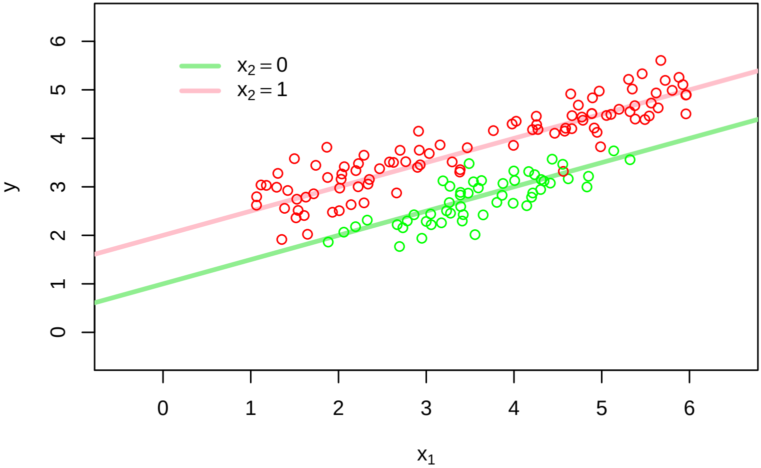
$$\begin{aligned} E[y|x_1] &= E[\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2|x_1] \\ &= \alpha_0 + \alpha_1 x_1 + \alpha_2 E[x_2|x_1] \\ &= \alpha_0 + \alpha_1 x_1 + \alpha_2 p(x_1) \end{aligned}$$

$E[y|x_1, x_2]$ linear in x_1, $E[y|x_1]$ nonlinear in x_1.

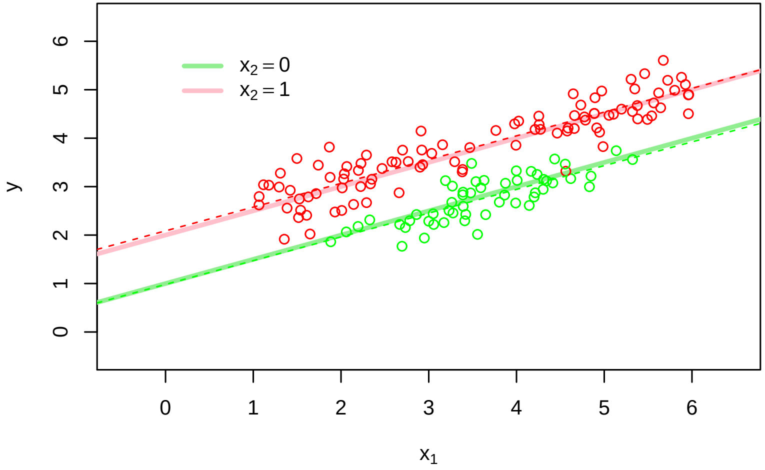
(This is because $E[x_2|x_1]$ is nonlinear in x_1).

The "marginal effect" of x_1 is *not* the same its conditional effect.

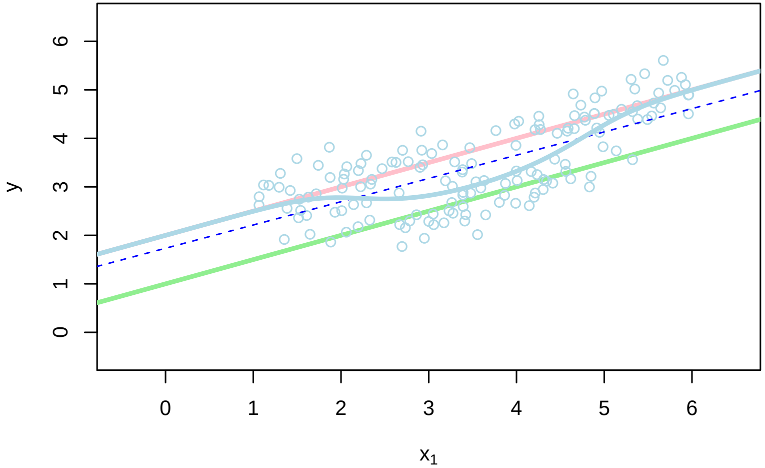
(This is because $E[x_2|x_1]$ depends on x_1).



$x_2 = 0$
$x_2 = 1$
y
x_1
0
1
2
3
4
5
6



$x_2 = 0$
$x_2 = 1$
y
x_1
0
1
2
3
4
5
6



y
0
1
2
3
4
5
6
0
1
2
3
4
5
6
x_1

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.

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.

In other words,

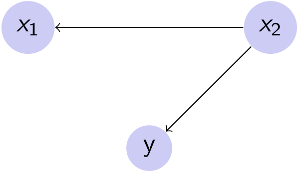
- the *conditional effect* of x_1 on y given x_2 is zero;
- x_1 and x_2 are correlated;
- the *conditional effect* of x_2 on y given x_1 is nonzero.
- the *marginal effect* of x_1 on y is nonzero.

Are marginal or conditional effects of the drug more relevant?

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

Graphical model

Drug example:



x_1 has no *causal* effect

x_1 has no *conditional* effect

x_1 has a *marginal* effect

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 $y \sim x_1$

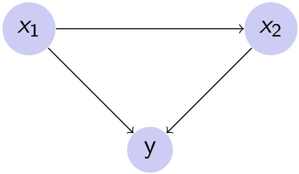
Suppose $\hat{\alpha}_1$ is smaller, less significant than $\hat{\beta}_1$.

Why might this be?

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

Graphical model

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

Changes to significance

The "statistical significance" of θ generally refers to its t-statistic,

$$t(\theta) = \frac{\hat{\theta}}{SE(\hat{\theta})},$$

and things derived from it (p-values, tests).

These things evaluate evidence that $\theta = 0$ versus $\theta \neq 0$.

If $\mathbf{x}_1$ and $\mathbf{x}_2$ are correlated in the sample, then

- $\hat{\alpha}_1$ and $\hat{\beta}_1$ are generally different;
- it should not be a surprise that $t(\alpha_1) \neq t(\beta_1)$.

If $\mathbf{x}_1$ and $\mathbf{x}_2$ are uncorrelated in the sample, then $\hat{\alpha}_1 = \hat{\beta}_1$.

Q: Can $t(\alpha_1) \neq t(\beta_1)$?
Q: Can adding x_2 to the model affect the significance of x_1?

Suppose we have data $\mathbf{y}$, $\mathbf{x}_1$, $\mathbf{x}_2$, where

- $\mathbf{x}_1$ and $\mathbf{x}_2$ are centered,
- $\mathbf{x}_1$ and $\mathbf{x}_2$ have zero sample correlation.

Let $t(\alpha_1)$ be the t-stat for x_1 from `lm( y ~ x1 + x2 )`.

Let $t(\beta_1)$ be the t-stat for x_1 from `lm(y ~ x1)`.

$\hat{\alpha}_1 = \hat{\beta}_1$. What about the t-statistics?

You should be able to show that in this case,

$$t(\alpha_1) = \frac{\mathbf{x}_1^T \mathbf{y}}{\sqrt{\hat{\sigma}_c^2 / SX_1X_1}} \quad , \quad t(\beta_1) = \frac{\mathbf{x}_1^T \mathbf{y}}{\sqrt{\hat{\sigma}_m^2 / SX_1X_1}}$$

where

- $\hat{\sigma}_c^2$ is the variance estimate from the conditional model;
- $\hat{\sigma}_m^2$ is the variance estimate from the marginal model.

Roughly speaking (ignoring the dof of the t-distribution),

- $\hat{\sigma}_c^2 < \hat{\sigma}_m^2 \Rightarrow$ adding x_2 increases the significance of x_1;
- $\hat{\sigma}_c^2 > \hat{\sigma}_m^2 \Rightarrow$ adding x_2 decreases the significance of x_1.

Case 1:

If

- $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are uncorrelated in the sample, and
- $\boldsymbol{x}_2$ has a nonzero effect (is correlated with $\boldsymbol{y}$),

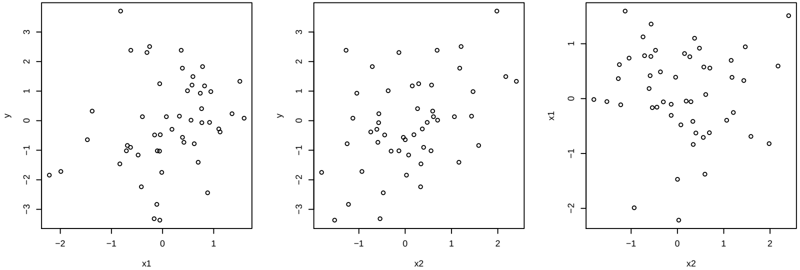
then typically $\hat{\sigma}_c^2 < \sigma_m^2$ and significance of x_1 is increased by adding x_2.

Why?

- Since $\boldsymbol{x}_2$ is correlated with $\boldsymbol{y}$, adding it reduces RSS.
- Lower RSS means lower variance estimate.

This holds as long as the improvement is sufficiently large.

Numerical example

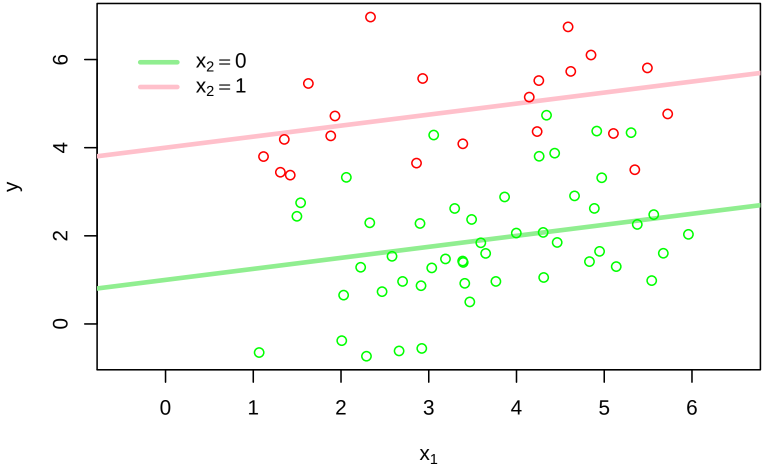


Numerical example

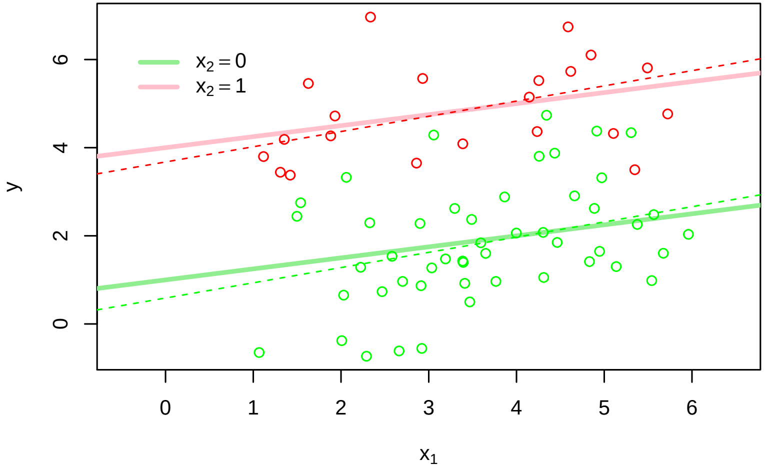
```
cat( cor(x1,y) , cor(x2,y) , cor(x1,x2) )

## 0.2202513 0.4320678 -0.03908718
```

```
summary(lm(y~x1))$coef

##               Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -0.1476833  0.2177266 -0.6782971 0.5008405
## x1           0.4107883  0.2625917  1.5643615 0.1243024
```

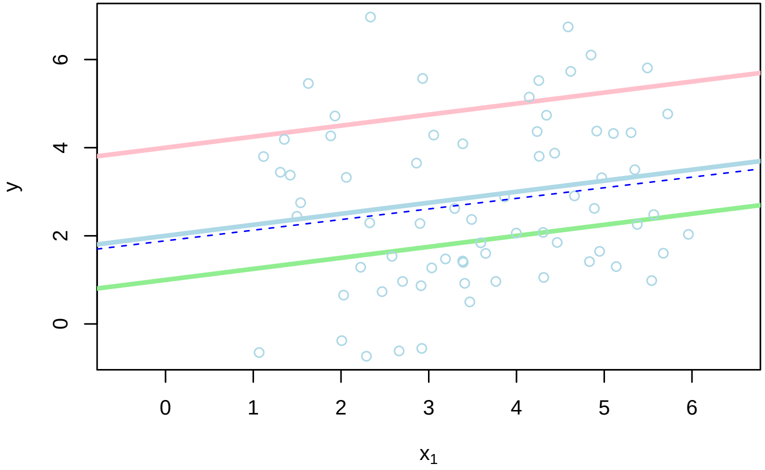
```
summary(lm(y~x1+x2))$coef

##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) -0.2337934  0.1978165 -1.181870 0.243202642
## x1           0.4429633  0.2368806  1.869985 0.067721601
## x2           0.7063892  0.2032777  3.474995 0.001109098
```



$x_2 = 0$
$x_2 = 1$
y
0
2
4
6
0
1
2
3
4
5
6
x_1



$x_2 = 0$
$x_2 = 1$
y
0
2
4
6
0
1
2
3
4
5
6
x_1



6
4
2
0
y
0
1
2
3
4
5
6
x_1

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

```
##               Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 0.5883570  0.4242843 1.386705 1.701286e-01
## x1          0.3453634  0.1066730 3.237590 1.875951e-03
## x2          3.0870588  0.3109921 9.926487 8.638400e-15
```

```r
sum(lm(y~x1+x2)$res^2)
```

```
## [1] 94.32418
```

```r
sum(lm(y~x1+x2)$res^2)/(n-3)
```

```
## [1] 1.407824
```

```r
summary(lm(y~x1))$coef
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 1.887218  0.6297176 2.996927 0.003804579
## x1          0.240653  0.1656194 1.453048 0.150809728
```

```r
sum(lm(y~x1)$res^2)
```

```
## [1] 233.0443
```

```r
sum(lm(y~x1)$res^2)/(n-2)
```

```
## [1] 3.427122
```

Decreasing significance

If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are *correlated*, then adding $\boldsymbol{x}_2$ can definitely decrease the significance of $\boldsymbol{x}_1$.

If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are *not correlated*, then it is more unusual for this to happen.

- adding variables always decreases RSS;
- adding variables will decrease $\hat{\sigma}^2$ unless the decrease in RSS is outweighed by the decrease in dof.

Simulation example

```r
n<-50 ; p<-20

X<-matrix(rnorm(n*p),n,p)

y<-X[,1]/2 + rnorm(n)
```

```r
summary(lm(y~X[,1]))$coef

##                Estimate Std. Error    t value   Pr(>|t|)
## (Intercept) -0.01859293  0.1391084 -0.1336578 0.89423196
## X[, 1]       0.45433113  0.1677734  2.7080051 0.00935132
```

```
summary(lm(y~X))$coef

##                Estimate Std. Error     t value   Pr(>|t|)
## (Intercept)  0.01242748  0.1771138  0.07016664 0.94454266
## X1           0.61180114  0.2618203  2.33672153 0.02656961
## X2          -0.25285253  0.1933493 -1.30774987 0.20123083
## X3          -0.18650060  0.2100119 -0.88804788 0.38182162
## X4           0.26432270  0.2101902  1.25754048 0.21859090
## X5           0.08152787  0.1642565  0.49634497 0.62339243
## X6           0.19944125  0.2028733  0.98308286 0.33369784
## X7          -0.02258492  0.2167142 -0.10421520 0.91771600
## X8          -0.02127171  0.1811366 -0.11743465 0.90732521
## X9          -0.03738690  0.1553097 -0.24072489 0.81146210
## X10          0.16309520  0.1723195  0.94647001 0.35172913
## X11          0.33881042  0.2128480  1.59179553 0.12227392
## X12          0.03595129  0.1940826  0.18523712 0.85433252
## X13          0.08023394  0.1836761  0.43682301 0.66547511
## X14          0.31933321  0.1674199  1.90737879 0.06641988
## X15          0.07463418  0.1565567  0.47672309 0.63713201
## X16         -0.08415989  0.1850508 -0.45479355 0.65264411
## X17         -0.06178351  0.1535279 -0.40242541 0.69032200
## X18          0.05473411  0.2144753  0.25520006 0.80036984
## X19          0.08365743  0.1728615  0.48395633 0.63205152
## X20         -0.09365810  0.1775308 -0.52755975 0.60181831
```

```r
fit1<-lm(y~X[,1]) ; fitfull<-lm(y~X)
```

```r
sum(fit1$res^2)

## [1] 45.76116

n-length(fit1$coef)

## [1] 48

summary(fit1)$sigma

## [1] 0.9764003
```

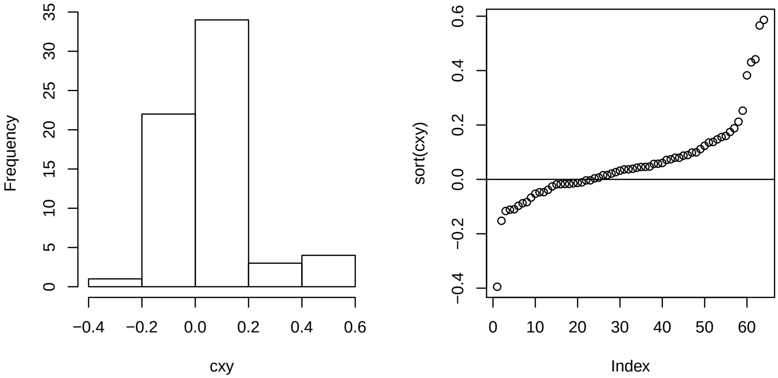
```r
sum(fitfull$res^2)

## [1] 31.02169

n-length(fitfull$coef)

## [1] 29

summary(fitfull)$sigma

## [1] 1.03427
```

Diabetes example:

$$
\begin{aligned}
y &= \text{diabetes progression} \\
x_1 &= \text{age} \\
x_1 &= \text{sex} \\
&\vdots
\end{aligned}
$$

```
dim(X)

## [1] 442  64

colnames(X)

##  [1] "age"     "sex"     "bmi"     "map"     "tc"      "ldl"     "hdl"
##  [8] "tch"     "ltg"     "glu"     "age^2"   "bmi^2"   "map^2"   "tc^2"
## [15] "ldl^2"   "hdl^2"   "tch^2"   "ltg^2"   "glu^2"   "age:sex" "age:bmi"
## [22] "age:map" "age:tc"  "age:ldl" "age:hdl" "age:tch" "age:ltg" "age:glu"
## [29] "sex:bmi" "sex:map" "sex:tc"  "sex:ldl" "sex:hdl" "sex:tch" "sex:ltg"
## [36] "sex:glu" "bmi:map" "bmi:tc"  "bmi:ldl" "bmi:hdl" "bmi:tch" "bmi:ltg"
## [43] "bmi:glu" "map:tc"  "map:ldl" "map:hdl" "map:tch" "map:ltg" "map:glu"
## [50] "tc:ldl"  "tc:hdl"  "tc:tch"  "tc:ltg"  "tc:glu"  "ldl:hdl" "ldl:tch"
## [57] "ldl:ltg" "ldl:glu" "hdl:tch" "hdl:ltg" "hdl:glu" "tch:ltg" "tch:glu"
## [64] "ltg:glu"
```

```
cxy<-apply( X, 2, function(x){ cor(x,y) } )
```



```r
fit_full<-lm(y ~ -1+ X)

summary(fit_full)$sigma

## [1] 0.6895559

summary(fit_full)$coef[1:25,]

##                Estimate  Std. Error     t value     Pr(>|t|)
## Xage        0.031329759  0.04041242  0.77525078 4.386762e-01
## Xsex       -0.165133830  0.04026257 -4.10142300 5.030866e-05
## Xbmi        0.284579388  0.05218759  5.45300911 8.962419e-08
## Xmap        0.211824072  0.04469026  4.73982583 3.033978e-06
## Xtc        -2.223376362 37.36674517 -0.05950147 9.525841e-01
## Xldl        1.870518158 32.84111892  0.05695659 9.546099e-01
## Xhdl        0.681333527 13.96347853  0.04879397 9.611093e-01
## Xtch        0.046287516  0.17013560  0.27206250 7.857225e-01
## Xltg        1.129254674 12.28457112  0.09192463 9.268066e-01
## Xglu        0.038762059  0.04342631  0.89259391 3.726426e-01
## Xage^2      0.041811619  0.04285347  0.97568807 3.298430e-01
## Xbmi^2      0.028320011  0.05137777  0.55121142 5.818144e-01
## Xmap^2     -0.005225396  0.04419948 -0.11822300 9.059538e-01
## Xtc^2       4.118988206  4.35455182  0.94590405 3.448018e-01
## Xldl^2      2.213266400  3.28551687  0.67364329 5.009498e-01
## Xhdl^2      1.069716832  0.98117020  1.09024594 2.762993e-01
## Xtch^2      0.477700198  0.37441687  1.27585115 2.027916e-01
## Xltg^2      0.896617287  1.06724113  0.84012625 4.013687e-01
## Xglu^2      0.070507611  0.05806088  1.21437378 2.253631e-01
## Xage:sex    0.091835722  0.04528227  2.02807226 4.325373e-02
## Xage:bmi   -0.011150367  0.04911505 -0.22702545 8.205267e-01
## Xage:map    0.011447848  0.04706844  0.24321705 8.079692e-01
## Xage:tc    -0.098144364  0.38067330 -0.25781783 7.966878e-01
## Xage:ldl    0.041560719  0.30505697  0.13623920 8.917047e-01
```

```r
summary(fit_full)$coef[26:64,]
```

```
##                Estimate Std. Error    t value   Pr(>|t|)
## Xage:tch  0.11424678 0.12974552  0.8805450 0.37912361
## Xage:ltg  0.07700475 0.13803294  0.5578722 0.57726187
## Xage:glu  0.03865170 0.04958159  0.7795575 0.43613916
## Xsex:bmi  0.03990987 0.04805484  0.8305067 0.40677662
## Xsex:map  0.05464750 0.04610680  1.1852373 0.23666821
## Xsex:tc   0.26782601 0.36438795  0.7350024 0.46279379
## Xsex:ldl -0.21793287 0.28927994 -0.7533632 0.45170064
## Xsex:hdl -0.07704431 0.16894070 -0.4560435 0.64862060
## Xsex:tch -0.08105418 0.12319658 -0.6579256 0.51098641
## Xsex:ltg -0.07350118 0.13971585 -0.5260762 0.59914380
## Xsex:glu  0.02826389 0.04543242  0.6221084 0.53424563
## Xbmi:map  0.09556772 0.05326008  1.7943593 0.07355479
## Xbmi:tc  -0.18656796 0.41202295 -0.4528096 0.65094561
## Xbmi:ldl  0.14919508 0.34607751  0.4311031 0.66663916
## Xbmi:hdl  0.07532162 0.20349410  0.3701415 0.71148434
## Xbmi:tch -0.02065841 0.14239485 -0.1450783 0.88472635
## Xbmi:ltg  0.07083188 0.15790951  0.4485599 0.65400617
## Xbmi:glu  0.01443987 0.05615752  0.2571315 0.79721716
## Xmap:tc   0.29543999 0.42086498  0.7019828 0.48312188
## Xmap:ldl -0.20182155 0.35427608 -0.5696731 0.56923783
## Xmap:hdl -0.11569484 0.19097491 -0.6058117 0.54500336
## Xmap:tch -0.03600725 0.12251026 -0.2939121 0.76898639
## Xmap:ltg -0.09561435 0.16776624 -0.5699260 0.56906643
## Xmap:glu -0.08244557 0.05632833 -1.4636608 0.14411772
## Xtc:ldl  -5.75296183 7.26125971 -0.7922815 0.42869348
## Xtc:hdl  -2.42874540 2.35431164 -1.0316159 0.30291167
## Xtc:tch  -1.36255310 1.08682047 -1.2537058 0.21072379
## Xtc:ltg  -2.34808672 8.12170784 -0.2891124 0.77265383
## Xtc:glu  -0.10889430 0.36731837 -0.2964575 0.76704351
## Xldl:hdl  1.63231706 1.95295045  0.8358210 0.40378360
```

Backwards selection

```
Xred<-X
while( any( summary(lm(y~-1+Xred))$coef[,4]>.05 ))
{
 fit_red<-lm(y~-1+Xred)

 pval<-summary(fit_red)$coef[,4]

 Xred<-Xred[, -which.max(pval) ]
}
```

```
colnames(Xred)

##  [1] "sex"     "bmi"     "map"     "tc"      "ldl"     "ltg"     "ltg^2"
##  [8] "glu^2"   "age:sex" "bmi:map" "tc:ltg"  "ldl:ltg" "hdl:ltg"
```

Warning! *p*-values, CIs from such a procedure do not have their usual properties.

```r
summary(fit_full)$sigma
```

```
## [1] 0.6895559
```

```r
summary(fit_red)$sigma
```

```
## [1] 0.6752136
```

```r
summary(fit_red)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## Xredsex     -0.15025922 0.03602612 -4.170841 3.674216e-05
## Xredbmi      0.30788650 0.03971589  7.752225 6.617185e-14
## Xredmap      0.19982453 0.03777414  5.289982 1.954453e-07
## Xredtc      -0.44478073 0.10561232 -4.211448 3.093842e-05
## Xredldl      0.32682618 0.09923618  3.293417 1.071763e-03
## Xredltg      0.57383796 0.05414551 10.598072 1.817738e-23
## Xredltg^2    0.30734819 0.10590616  2.902081 3.897918e-03
## Xredglu^2    0.08226810 0.03331794  2.469183 1.393070e-02
## Xredage:sex  0.13100570 0.03296684  3.973863 8.294843e-05
## Xredbmi:map  0.08698595 0.03372885  2.578978 1.024140e-02
## Xredtc:ltg  -0.45085520 0.15780773 -2.856991 4.484764e-03
## Xredldl:ltg  0.37996802 0.12363217  3.073375 2.251513e-03
## Xredhdl:ltg  0.16662767 0.06322931  2.635292 8.711021e-03
```

```r
sum(fit_full$res^2)

## [1] 179.7342

length(y)-length(fit_full$coef)

## [1] 378
```

```r
sum(fit_red$res^2)

## [1] 195.5869

length(y)-length(fit_red$coef)

## [1] 429
```