Multilevel Structure

true

2022-09-06

Contents

1 Introduction

Associations between two variables can be *very different* (or even *reversed*) depending upon whether or not the analysis is "aware" of the grouped, nested, or clustered nature of the data. In multilevel analysis, the groups are often neighborhoods, communities, or even different countries.

A model that is "aware" of the clustered nature of the data may provide very different–likely better–substantive conclusions than a model that is not aware of the clustered nature of the data.

2 Set a Random Seed For Replicability

```
set.seed(1080) # random seed
```

3 Call The Libraries

```
library(ggplot2) # beautiful graphs

library(lme4) # MLM

library(sjPlot) # nice tables for MLM
```

4 Simulate Some Data

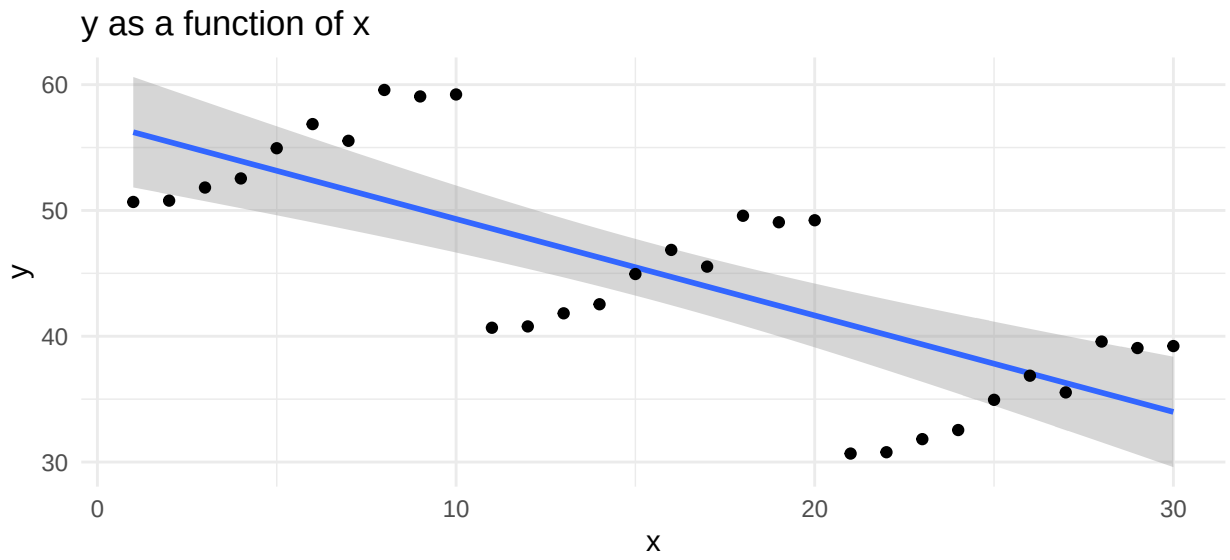
```
e <- rnorm(10, 0, 1) # error

# group 1

group1 <- rep(1, 10)

x1 <- seq(1,10)

y1 <- 50 + 1 * x1 + e

# group 2

group2 <- rep(2, 10)

x2 <- seq(11, 20)

y2 <- 30 + 1 * x2 + e

# group 3

group3 <- rep(3, 10)

x3 <- seq(21, 30)

y3 <- 10 + 1 * x3 + e

# combine into a dataframe

x <- c(x1, x2, x3)

y <- c(y1, y2, y3)

group <- factor(c(group1, group2, group3))

mydata <- data.frame(x, y, group)
```

5 Graphs

5.1 A "Naive" Graph

This "naive" graph is unaware of the grouped nature of the data.

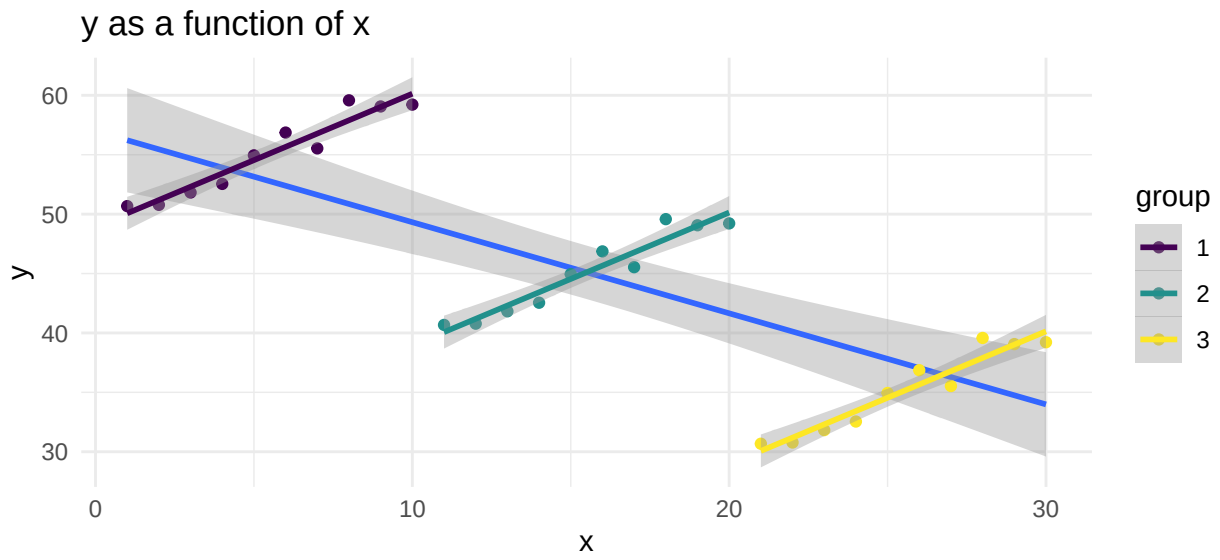
```
library(ggplot2)

p0 <- ggplot(mydata,
             aes(x = x,
                 y = y)) +
  geom_smooth(method = "lm") +
  labs(title = "y as a function of x") +
  theme_minimal()

p0 + geom_point() # replay and add points
```



5.2 An "Aware" Graph

This "aware" graph is aware of the grouped nature of the data.

```
p0 +
  geom_point(aes(color = group)) + # points with group color
  geom_smooth(aes(color = group), # smoothers with group color
              method = "lm") +
  scale_color_viridis_d()
```



6 Regressions

6.1 A "Naive" OLS Analysis

The OLS model with only x as a covariate is not aware of the grouped structure of the data, and the coefficient for x reflects this.

```
myOLS <- lm(y ~ x, data = mydata)

sjPlot::tab_model(myOLS,
                  show.se = TRUE,
                  show.ci = FALSE,
                  show.stat = TRUE)
```

y

Predictors

Estimates

std. Error

Statistic

p

(Intercept)

56.99

2.25

25.29

<0.001

x

-0.77

0.13

-6.04

<0.001

Observations

30

R2 / R2 adjusted

0.566 / 0.550

6.2 An "Aware" MLM Analysis

The multilevel model is aware of the grouped structure of the data, and the coefficient for x reflects this.

```
myMLM <- lmer(y ~ x + (1 | group), data = mydata)

sjPlot::tab_model(myMLM,
                  show.se = TRUE,
                  show.ci = FALSE,
                  show.stat = TRUE)
```

y

Predictors

Estimates

std. Error

Statistic

p

(Intercept)

27.82

12.25

2.27

0.032

x

1.11

0.06

17.87

<0.001

Random Effects

2

0.96

00 group

447.52

ICC

1.00

N group

3

Observations

30

Marginal R2 / Conditional R2

0.177 / 0.998

7 A Thought Experiment

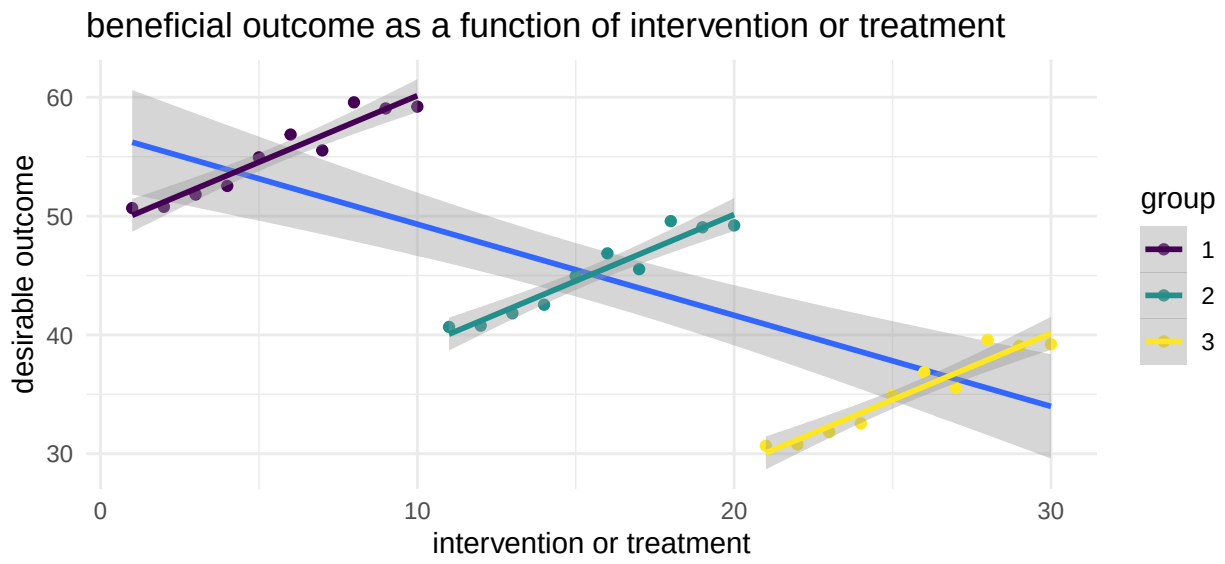
When might a situation like this arise in practice? This is surprisingly difficult to think through.

Imagine that x is a protective factor, or an intervention or treatment. Imagine that y is a desirable outcome, like improved mental health or psychological well being.

Now imagine that community members provide more of the protective factor or more of the intervention in situations where there are lower levels of the desirable outcome. If we think about it, this is a very plausible situation.

> A naive analysis that was unaware of the grouped nature of the data would therefore misconstrue the results, suggesting that the intervention was harmful, when it was in fact helpful.

```
p0 +
  geom_point(aes(color = group)) + # points with group color
  geom_smooth(aes(color = group), # smoothers with group color
              method = "lm") +
  labs(title = "beneficial outcome as a function of intervention or treatment",
       x = "intervention or treatment",
       y = "desirable outcome") +
  scale_color_viridis_d()
```



These data are constructed to provide this kind of extreme example, but it easy to see how multilevel thinking, and multilevel analysis may provide better answers than we would get if we ignored the grouped nature of the data.