Causal inference for experiments

Frank Edwards

Causal inference in experiments

Imagine a simple experiment where we evaluate student math test scores after exposure to supplemental tutoring.

Let y_i be a student's math test score

Let $x_i \in \{0, 1\}$ where one indicates a student received tutoring, and zero indicates no tutoring.

We would like to know the *causal effect* of tutoring on math scores.

The individual causal effect

Formally, we define the individual causal effect as $y_i(x = 1) - y_i(x = 0)$

For our example, this is the difference in test scores for student i when exposed to tutoring and not exposed to tutoring.

The fundamental problem of causal inference

We seek to learn what would have happened to unit *i*, which was exposed to treatment, if it had not been (or vice versa).

By definition, the *individual causal effect* cannot be identified.

The potential outcomes framework

Assume student *i* received tutoring, and had a test score of 82. We'd like to know what this student's score would have been under the *counterfactual* scenario of not receiving tutoring (but we lack a time machine).

For unit *i*, we only observe $y_i(x = 1)$ *OR* $y_i(x = 0)$, never both. The unobserved value of *y* is a *potential outcome*, or *counterfactual*.

Three kinds of treatment effects

An *individual treatment effect* is the causal effect of a treatment *x* on outcome *y* for a particular unit *i*. This is rarely identifiable.

A *sample average treatment effect* is the average of the causal effect of a treatment *x* on outcome *y* across units in a sample. This is straightforward to identify.

A *population average treatment effect* is the average of the causal effect of a treatment *x* on outcome *y* across a population. This is possible, if difficult, to identify.

Average treatment effects

If we have two groups of students, one exposed to tutoring and one not exposed to tutoring, we can easily compute the difference in average test scores as:

$\frac{1}{n}\sum_{i=1}^{n} y_i(x = 1) - \frac{1}{n}\sum_{i=1}^{n} y_i(x = 0)$

Under what conditions would this difference in means be equal to the *sample average treatment effect*?

Counterfactuals and comparison groups

For our difference in means to be equal to the SATE, each group must represent the *counterfactual* state for the other.

This holds (on average) when the two groups are identical on all important characteristics.

Making two (or more) identical groups

Simple random sampling will, *on average*, generate groups that are identical on measurable and unmeasurable traits.

There are other methods, like *propensity score matching* that seek to create groups that are statistically similar, but don't come with the assurances of balance on unmeasurables that come with randomization.

Putting it all together

Randomization of exposure to treatment ensures (assymptotically) identical groups.

While we still can't observe $y_i(x = 1) - y_i(x = 0)$, we now can compute $\frac{1}{n}\sum_{i=1}^{n} y_i(x = 1)$ and $\frac{1}{n}\sum_{i=1}^{n} y_i(x = 0)$, where each group (treatment, control), acts as the counterfactual for the other group.

Randomization and validity

Randomization (assuming a sufficiently large sample) insures that we can estimate an *unbiased* estimate of the sample average treatment effect (*Internal validity*).

Under which conditions would the sample average treatment effect also inform us about a *population* average treatment effect (*external validity*)?

Formal assumptions: Ignorability

Treatment assignment is not associated with the potential outcome:

$$x \perp y(x = 1), y(x = 0)$$

Having a greater or lesser likelihood to benefit from a treatment cannot be associated with assignment to treatment.

Balance

Randomization in general insures that, on average, we will have *balance* on all measured and unmeasured characteristics, including potential outcomes, across groups.

BUT we never know if any particular sample is balanced! It is critical to check as part of your analysis.

Stable unit treatment value assumption

- The potential outcome for unit *i* depends *only* on the treatment
    - No contamination or spillover effects
- The treatment is well defined and identical across units.

Compliance, intent to treat, missing data

Ordinary analysis assumes perfect compliance and complete data, but this is rarely the case.

Instead, we often conceptualize group assignment as *intent to treat*, with compliance with treatment requiring additional measurement.

Differential non-response can bias results even if we randomize! Always conduct missing data analyses and impute as needed.

Regression and experiments

The example

Randomized longitudinal experiment on the effects of guaranteed income ($500 per month for twelve months) on child and family well-being.

Families with prior low-level contact with New York Office of Children and Family Services in three counties were offered the opportunity to enroll. 50 from each county were randomized into treatment, and 100 from each county were randomized into control.

Measurement

Among other constructs, we measured *parental stress* with an 18-item instrument. Let's see if the treatment z had an effect on our measure of parental stress y.

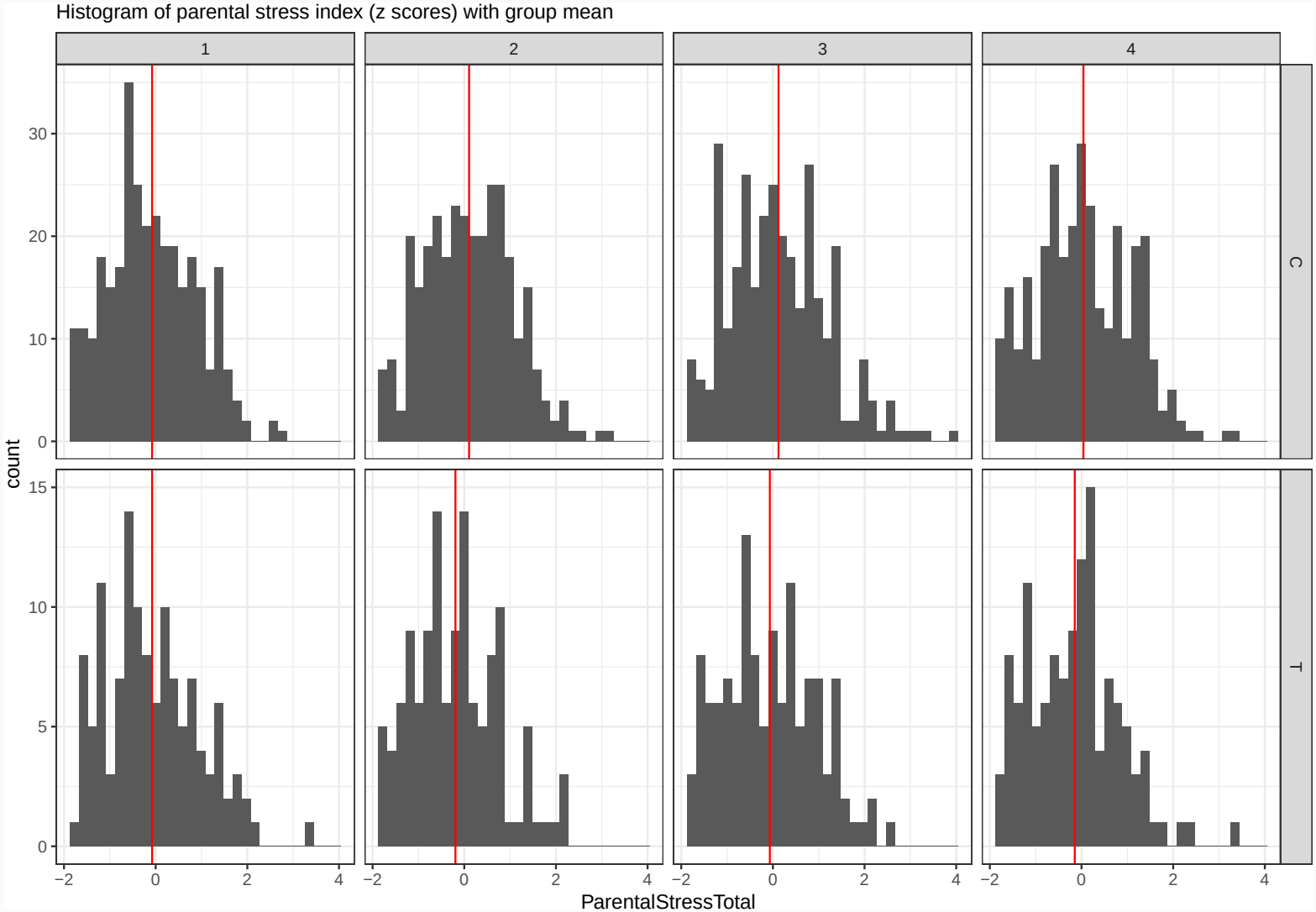
Identical surveys were collected at 4 periods: pre-treatment, 6 months after beginning treatment, 12 months after beginning treatment, and 18 months after beginning treatment.

Potential outcomes in this example

For parent i, we can describe two potential outcomes for parental stress: $y(z = 1)$ and $y(z = 0)$.

While we can't identify the *individual causal effect*, because we've randomized treatment (within blocks), we can assume that the treatment and control groups are effectively identical, and we can identify the *sample average treatment effect*.

The data


Histogram of parental stress index (z scores) with group mean
1
2
3
4
C
T
count
30
20
10
0
15
10
5
0
−2
0
2
4
ParentalStressTotal

The data

```
dat %>%
    group_by(GROUP, Wave) %>%
    summarize(PSS_mn = mean(ParentalStressTotal))
```

```
## # A tibble: 8 x 3
## # Groups:   GROUP [2]
##   GROUP  Wave  PSS_mn
##   <chr> <dbl>   <dbl>
## 1 C         1 -0.0780
## 2 C         2  0.105
## 3 C         3  0.123
## 4 C         4  0.0424
## 5 T         1 -0.0770
## 6 T         2 -0.193
## 7 T         3 -0.0667
## 8 T         4 -0.147
```

Setting up a simple regression model

Let's model parental stress as a function of the guaranteed income treatment only for the treatment period (Waves 2 and 3) using this model

$$y_i = \beta_0 + \beta_1 z_i + \varepsilon_{it}$$

Setting up a simple regression model

Let's model parental stress as a function of the guaranteed income treatment only for the treatment period (Waves 2 and 3) using this model

$y_i = \beta_0 + \beta_1 z_i + \varepsilon_{it}$

So $E(y_i|z_i = 0) = \beta_0$ and $E(y_i|z_i = 1) = \beta_0 + \beta_1$

Setting up a simple regression model

Let's model parental stress as a function of the guaranteed income treatment only for the treatment period (Waves 2 and 3) using this model

$y_i = \beta_0 + \beta_1 z_i + \varepsilon_{it}$

So $E(y_i|z_i = 0) = \beta_0$ and $E(y_i|z_i = 1) = \beta_0 + \beta_1$

Conveniently, the SATE is $E(y_i|z_i = 1) - E(y_i|z_i = 0) = \beta_1$

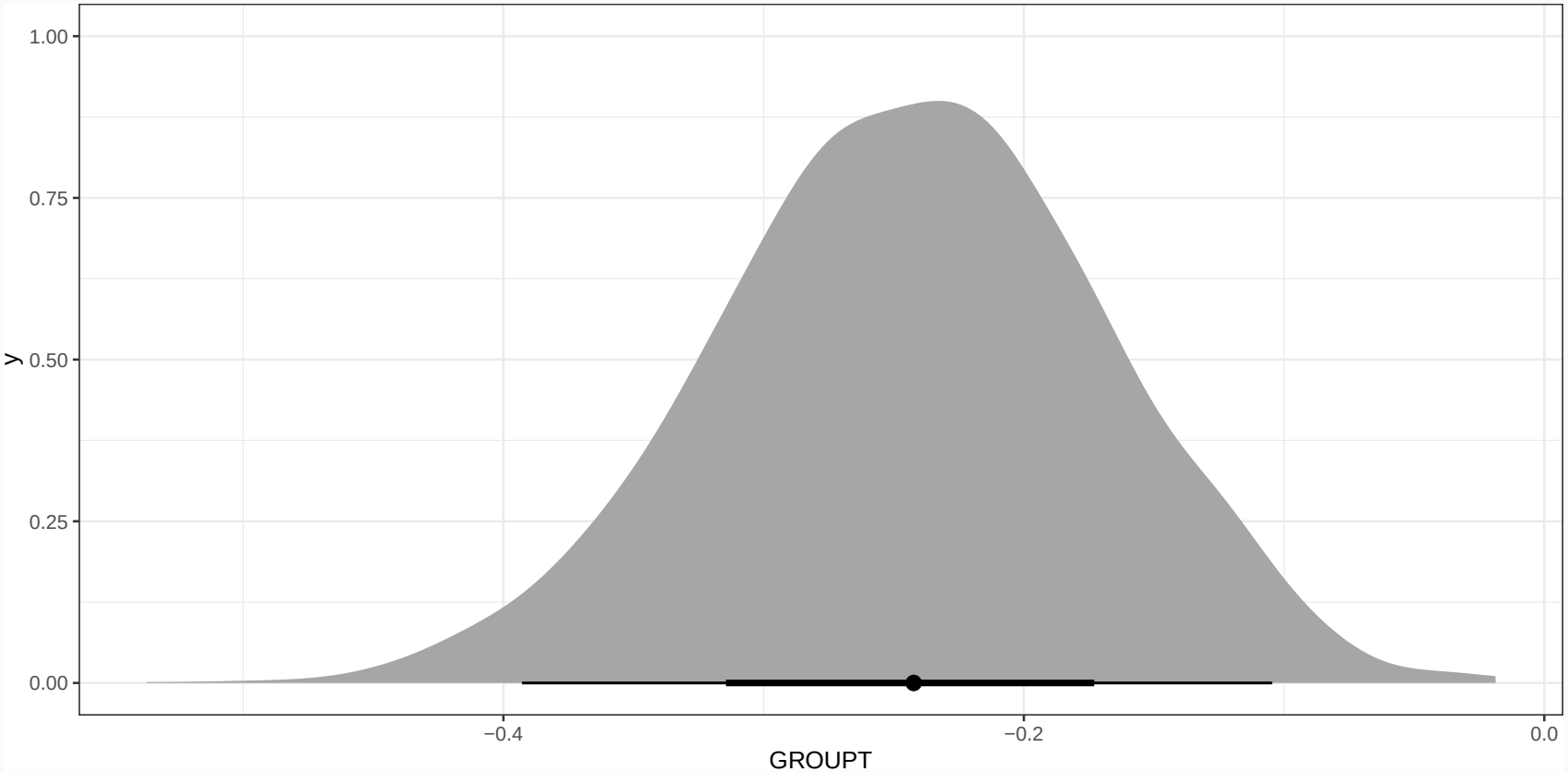
Estimating our model

```
m0 <- stan_glm(ParentalStressTotal ~ GROUP, data = dat %>%
    filter(Wave == 2 | Wave == 3), family = "gaussian", refresh = 0)

m0
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      ParentalStressTotal ~ GROUP
##  observations: 870
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept)  0.1    0.0
## GROUPT      -0.2    0.1
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Posterior inference: Beta 1

```
m0_post <- data.frame(m0)
ggplot(m0_post, aes(x = GROUPT)) + stat_slabinterval()
```



Incorporating the longitudinal design

Let's adjust our model to incorporate the time components of the design, now indicating measurement for individual i at time t

$$y_{it} = \beta_0 + \beta_1 z_i + \beta_2 t + \beta_3 z_i \times t + \varepsilon_{it}$$

Defining the average treatment effect

Because we are going to treat time as categorical, the model will actually look like this:

$$y_{it} = \beta_0 + \beta_1 z_i + \beta_2(t = 1) + \beta_3(t = 2) + \beta_4(t = 3) + \beta_5(t = 4) + \beta_6 z_i \times (t = 1) + \beta_7 z_i \times (t = 2) + \beta_8 z_i \times (t = 3) + \beta_9 z_i \times (t = 4) + \varepsilon_{it}$$

For wave 2 (6 months into treatment), expected values for treatment and control will be

$$E(y|z = 1, t = 2) = \beta_0 + \beta_1 + \beta_3 + \beta_7$$

$$E(y|z = 0, t = 2) = \beta_0 + \beta_3$$

So the SATE for wave 2 is $E(y|z = 1, t = 2) - E(y|z = 0, t = 2) = \beta_1 + \beta_7$

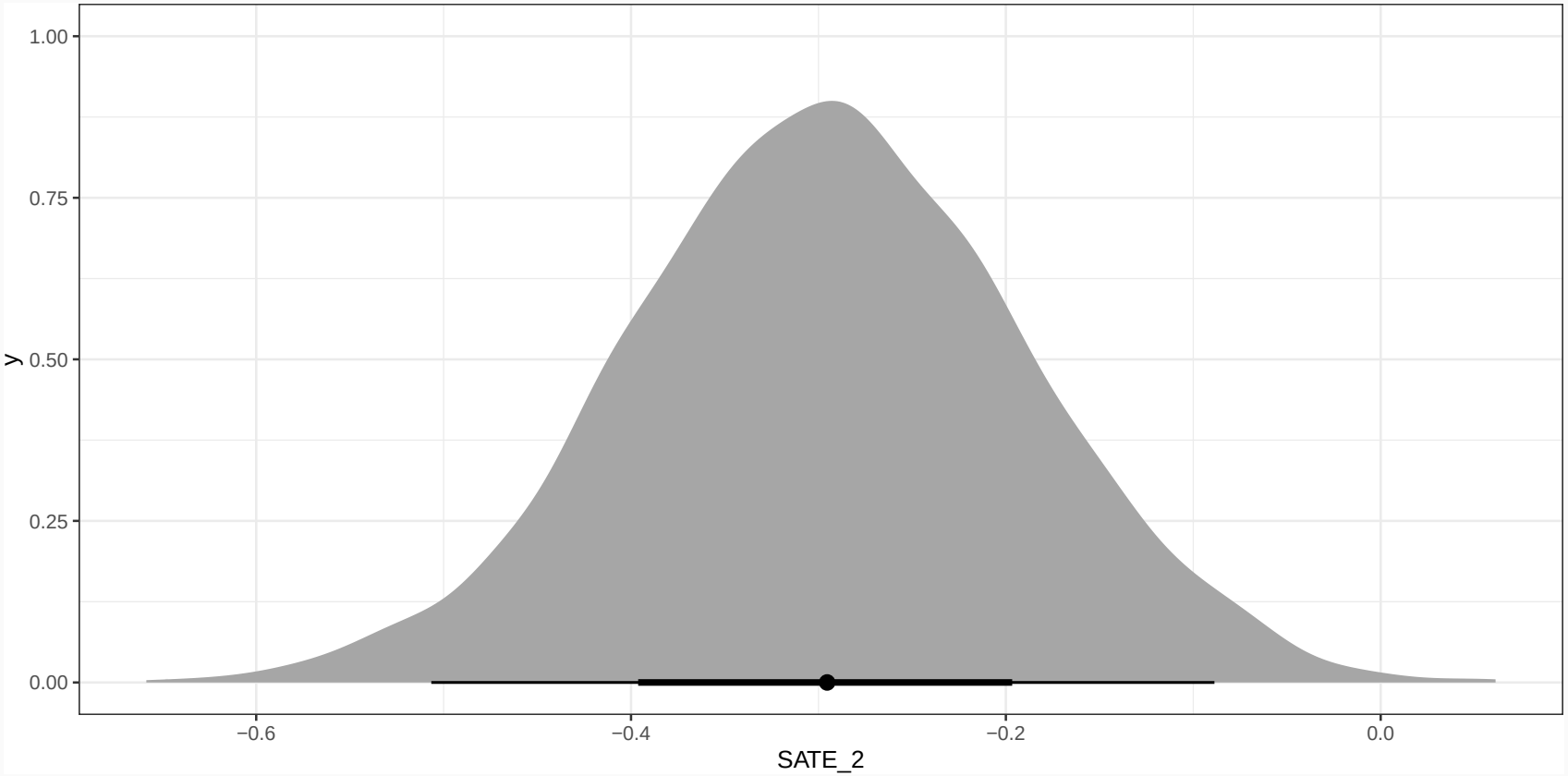
Estimating the model

```r
m1 <- stan_glm(ParentalStressTotal ~ GROUP * factor(Wave), data = dat, family = "gaussian",
    refresh = 0)

m1
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      ParentalStressTotal ~ GROUP * factor(Wave)
##  observations: 1740
##  predictors:   8
## ------
##                      Median MAD_SD
## (Intercept)          -0.1    0.1
## GROUPT                0.0    0.1
## factor(Wave)2         0.2    0.1
## factor(Wave)3         0.2    0.1
## factor(Wave)4         0.1    0.1
## GROUPT:factor(Wave)2 -0.3    0.1
## GROUPT:factor(Wave)3 -0.2    0.2
## GROUPT:factor(Wave)4 -0.2    0.2
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Posterior inference for the SATE, wave 2

```
m1_post <- data.frame(m1)
m1_post <- m1_post %>%
    mutate(SATE_2 = GROUPT + GROUPT.factor.Wave.2)

ggplot(m1_post, aes(x = SATE_2)) + stat_slabinterval()
```

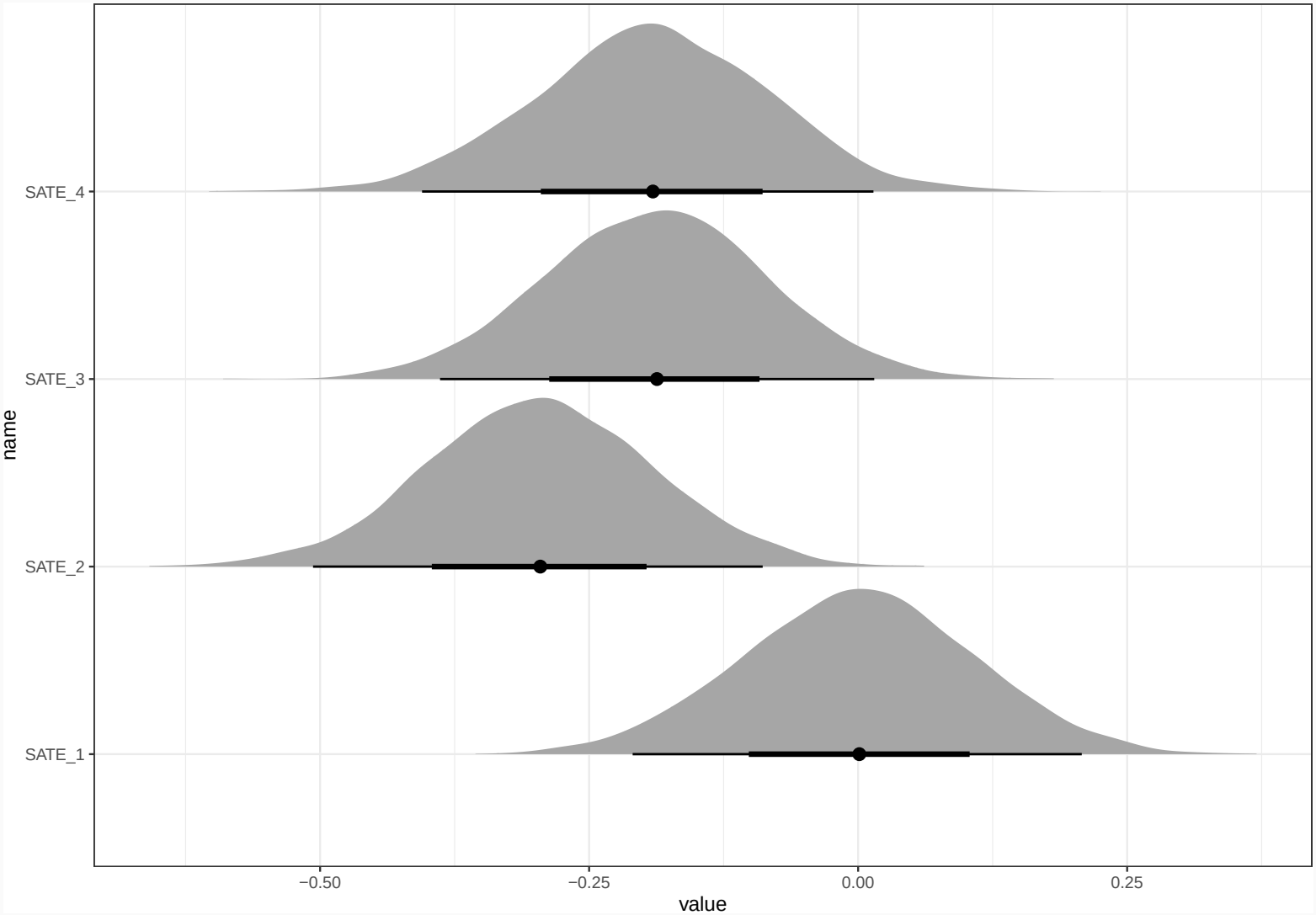


Let’s look at all waves

```
m1_post <- data.frame(m1) %>%
    mutate(SATE_1 = GROUPT, SATE_2 = GROUPT + GROUPT.factor.Wave.2, SATE_3 = GROUPT +
        GROUPT.factor.Wave.3, SATE_4 = GROUPT + GROUPT.factor.Wave.4) %>%
    select(SATE_1:SATE_4) %>%
    pivot_longer(cols = c(SATE_1, SATE_2, SATE_3, SATE_4))  # pivot to ggplot
```

Visualized

```
ggplot(m1_post, aes(y = name, x = value)) + stat_slabinterval()
```



Adjusting for pre-treatment differences

Our study took place in three counties with differing socio-economic contexts. Design features of the study, like *blocks* or *strata* for sampling, should be reflected in our analysis to avoid bias.

We can also adjust for pre-treatment characteristics if we are concerned about balance or efficiency.

It is almost always a *BAD* idea to include post-treatment measures in a causal model

Our new model:

$$y_{it} = \beta_0 + \beta_1 z_i + \beta_2 t + \beta_3 z_i \times t + \beta_4 County_i + \varepsilon_{it}$$

Estimation

```
m2 <- stan_glm(ParentalStressTotal ~ GROUP * factor(Wave) + County, data = dat, family = "gaussian",
    refresh = 0)

m2
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      ParentalStressTotal ~ GROUP * factor(Wave) + County
##  observations: 1740
##  predictors:   10
## ------
##                           Median MAD_SD
## (Intercept)               -0.1    0.1
## GROUPT                     0.0    0.1
## factor(Wave)2              0.2    0.1
## factor(Wave)3              0.2    0.1
## factor(Wave)4              0.1    0.1
## CountyOnondaga County      0.0    0.1
## CountyWestchester County   0.0    0.1
## GROUPT:factor(Wave)2      -0.3    0.2
## GROUPT:factor(Wave)3      -0.2    0.1
## GROUPT:factor(Wave)4      -0.2    0.1
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
```

Introducing multilevel modeling

We have an additional issue that needs to be addressed: we are repeating measurement on the same individuals many times, violating the assumption that our errors are independent and identically distributed (iid).

To fix this, we can add an extra intercept to our model, one for each individual.

This functions like an extra error term, adjusting for time-stable differences in reported stress across individuals.

Model specification

Let's take our county adjusted model and add *random intercepts* into the model for each individual.

$$y_{it} = \beta_0 + \beta_1 z_i + \beta_2 t + \beta_3 z_i \times t + \beta_4 County_i + \delta_i + \varepsilon_{it}$$

We made one small addition to the linear predictor, and we'll make one extra assumption to the stochastic component of the model

$$\varepsilon \sim N(0, \sigma^2)$$

$$\delta \sim N(0, \sigma_\delta^2)$$

This adjusts for repeated measurement of individuals by adding in extra error to the model.

Estimation

For multi-level models, we'll use **`stan_glmer`** (generalized linear mixed effects regression).

```
m3<-stan_glmer(ParentalStressTotal ~
                 GROUP * factor(Wave) + County +
                 (1|PIN),#this means estimate a random intercept for each PIN
               data = dat,
               family = "gaussian",
               refresh = 0)
```

Model output

```
m3
```

```
## stan_glmer
##  family:       gaussian [identity]
##  formula:      ParentalStressTotal ~ GROUP * factor(Wave) + County + (1 | PIN)
##  observations: 1740
## ------
##                           Median MAD_SD
## (Intercept)               -0.1    0.1
## GROUPT                     0.0    0.1
## factor(Wave)2              0.2    0.0
## factor(Wave)3              0.2    0.0
## factor(Wave)4              0.1    0.0
## CountyOnondaga County      0.0    0.1
## CountyWestchester County   0.0    0.1
## GROUPT:factor(Wave)2      -0.3    0.1
## GROUPT:factor(Wave)3      -0.2    0.1
## GROUPT:factor(Wave)4      -0.2    0.1
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.6    0.0
##
## Error terms:
##  Groups   Name        Std.Dev.
##  PIN      (Intercept) 0.79
##  Residual             0.61
## Num. levels: PIN 435
##
```

Compare to model 2

```
m2
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      ParentalStressTotal ~ GROUP * factor(Wave) + County
##  observations: 1740
##  predictors:   10
## ------
##                           Median MAD_SD
## (Intercept)               -0.1    0.1
## GROUPT                     0.0    0.1
## factor(Wave)2              0.2    0.1
## factor(Wave)3              0.2    0.1
## factor(Wave)4              0.1    0.1
## CountyOnondaga County      0.0    0.1
## CountyWestchester County   0.0    0.1
## GROUPT:factor(Wave)2      -0.3    0.2
## GROUPT:factor(Wave)3      -0.2    0.1
## GROUPT:factor(Wave)4      -0.2    0.1
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Posterior inference

Our inference approach can remain unchanged (check the algebra though!)

```
m3_post <- data.frame(m3) %>%
    mutate(SATE_1 = GROUPT, SATE_2 = GROUPT + GROUPT.factor.Wave.2, SATE_3 = GROUPT +
        GROUPT.factor.Wave.3, SATE_4 = GROUPT + GROUPT.factor.Wave.4) %>%
    select(SATE_1:SATE_4) %>%
    pivot_longer(cols = c(SATE_1, SATE_2, SATE_3, SATE_4))
```

Visualized

```
ggplot(m3_post, aes(y = name, x = value)) + stat_slabinterval()
```