

Randomized controlled trial

Researchers from the Jameel Poverty Action Lab (J-PAL) conducted a randomized controlled trial (RCT) of a technical training program in two villages in Ecuador. The program was designed to increase the income of participants by teaching them R (not really—it's a fake program, so just pretend).

Two villages were chosen at random, and one of the two was selected to be treatment group. Every individual over the age of 18 received the program (and assume that they all complied and attended). Nobody in the control group received the program (again, assume perfect compliance).

Your colleague attempted to measure the causal effect of this experimental training program on incomes. They conducted some statistical analysis in R, but they forgot to interpret anything in the document, and now they've moved to a different office!

You have access to the original data, which contains these columns:

Variable name	Description
ID	Participant ID number
Age	Participant age
Sex	Participant sex
pre_income	Monthly income before intervention (in USD)
post_income	Monthly income after intervention (in USD)
experiment_group	Indicator marking if participant was assigned to treatment or control

Given the information provided below, interpret the results from this analysis, as well as any assumption checks or tests your colleague included. Did this program have have an effect on wages? How much? Is it significant?

```
library(tidyverse)
library(broom)

village_rct <- read_csv("village_experiment.csv")

head(village_rct)
```

ID	Age	Sex	pre_income	post_income	experiment_group
001	45	Male	893	1009	Treatment
002	39	Female	871	707	Treatment
003	26	Male	912	1094	Treatment
004	22	Male	834	751	Treatment
005	55	Male	841	930	Treatment
006	33	Female	862	1116	Treatment

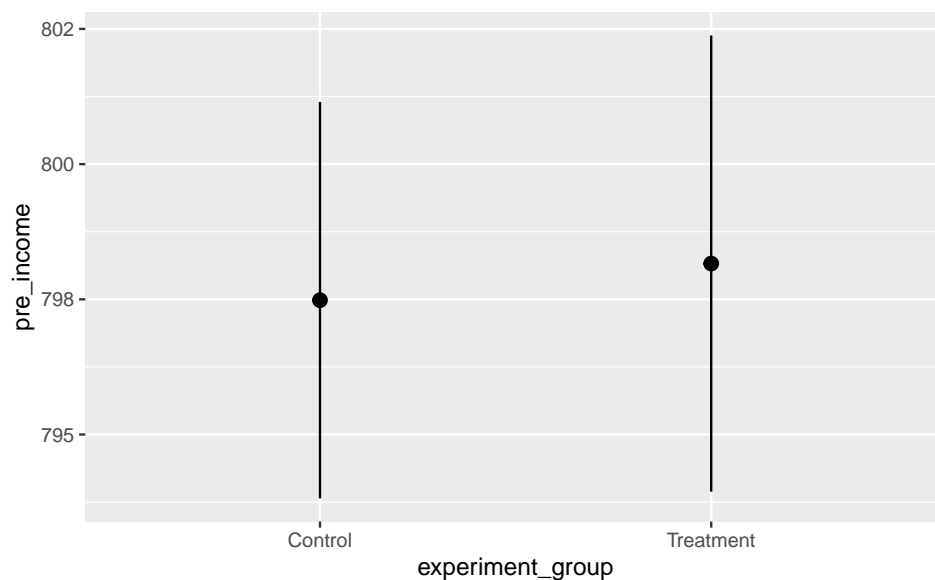
↓ 1: What's going on here? ↓

```
village_rct %>%
  group_by(experiment_group) %>%
  summarize(avg_age = mean(Age),
            prop_male = mean(Sex == "Male"),
            avg_pre_income = mean(pre_income),
            avg_post_income = mean(post_income))
```

experiment_group	avg_age	prop_male	avg_pre_income	avg_post_income
Control	54.0	0.592	797	797
Treatment	52.5	0.606	798	887

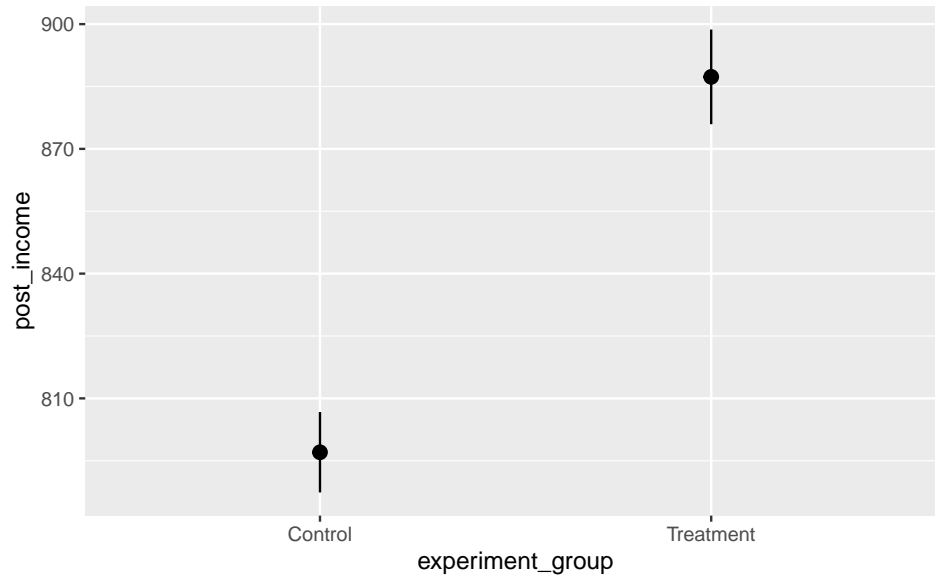
↓ 2: What's going on here? ↓

```
ggplot(village_rct, aes(x = experiment_group, y = pre_income)) +
  stat_summary(fun.data = "mean_se")
```



↓ 3: What's going on here? ↓

```
ggplot(village_rct, aes(x = experiment_group, y = post_income)) +  
  stat_summary(fun.data = "mean_se")
```



↓ 4: What's going on here? ↓

```
model1 <- lm(post_income ~ experiment_group, data = village_rct)  
tidy(model1)
```

term	estimate	std.error	statistic	p.value
(Intercept)	797.0	9.76	81.65	0
experiment_groupTreatment	90.3	14.89	6.06	0

```
glance(model1)
```

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
0.029	0.028	261	36.8	0	1	-8742	17489	17505	85052630	1250	1252