

9.5 observational studies

Chris Parrish

July 3, 2016

Contents

9.5 observational studies	1
randomized experiments	1
data	1
model	1
fit	2
figure 9.9	2
figure 9.12	3

9.5 observational studies

reference:

- ARM chapter 09, github

```
library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
library(ggplot2)
```

randomized experiments

data

```
### Data
# electric_gradeX_supp.data.R, where X = 1, 2, 3, 4
```

model

electric_supp.stan

```
data {
  int<lower=0> N;
  vector[N] post_test;
  vector[N] supp;
  vector[N] pre_test;
}
parameters {
  vector[3] beta;
  real<lower=0> sigma;
```

```

}
model {
  post_test ~ normal(beta[1] + beta[2] * supp + beta[3] * pre_test, sigma);
}

```

fit

```

### Model: post_test ~ supp + pretest
### Figures
beta1 <- beta2 <- beta3 <- rep(NA, 4)
sd2 <- rep(NA, 4)
# empty data frames
supp_effect.ggdf <- data.frame(c(), c(), c()) # Figure 9.9
prepost.ggdf <- data.frame(c(), c(), c(), c(), c(), c(), c()) # Figure 9.12

for (i in 1:4) {
  source(paste("electric_grade", i, "_supp.data.R", sep = ""))
  data.list <- c("N", "post_test", "pre_test", "supp")
  sf <- stan(file='electric_supp.stan', data=data.list,
            iter=1000, chains=4)
  beta.post <- extract(sf, "beta")$beta
  beta1 <- mean(beta.post[,1])
  beta2 <- mean(beta.post[,2])
  beta3 <- mean(beta.post[,3])
  sd2 <- sd(beta.post[,2])
  supp_effect.ggdf <- rbind(supp_effect.ggdf,
                           data.frame(est = beta2, se = sd2, grade = i))
  prepost.ggdf <- rbind(prepost.ggdf,
                       data.frame(post_test, pre_test, grade = i, supp,
                                   beta1, beta2, beta3))
}

```

```

## Warning: There were 3 divergent transitions after warmup. Increasing
## adapt_delta above 0.8 may help.

## Warning: Examine the pairs() plot to diagnose sampling problems

```

figure 9.9

```

## Figure 9.9
p1 <- ggplot(supp_effect.ggdf, aes(x = 5 - grade, y = est)) +
  geom_pointrange(aes(ymin = est - se, ymax = est + se), size = 0.8) +
  geom_pointrange(aes(ymin = est - 2 * se, ymax = est + 2 * se), size = 0.3) +
  geom_hline(aes(yintercept = 0), linetype = "dashed") +
  scale_x_continuous("Subpopulation", breaks = seq(1,4),
                    labels = paste("Grade", seq(4,1,-1))) +
  scale_y_continuous("", breaks = c(0, 5, 10)) +
  coord_flip() +
  ggtitle("Estimated effect of supplement,\ncompared to replacement")
print(p1)

```

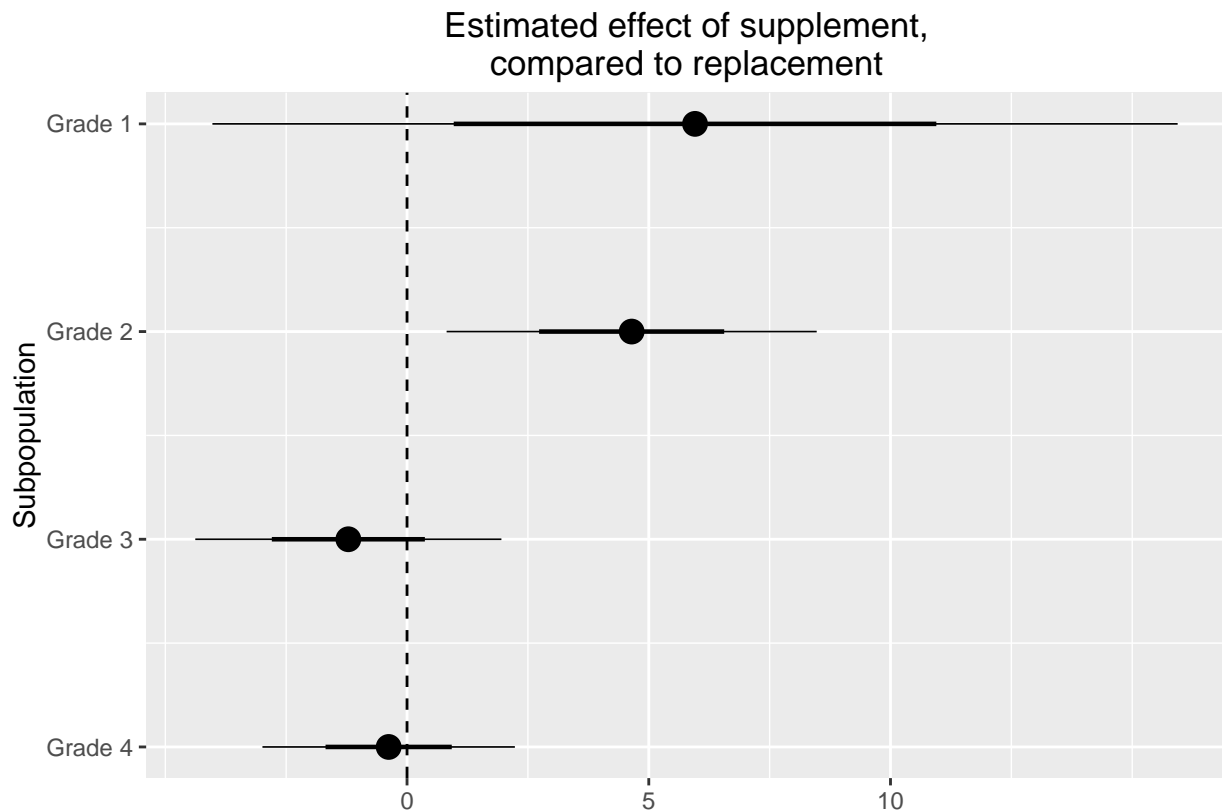


figure 9.12

```
## Figure 9.12
prepost.ggdf$supp <- factor(prepost.ggdf$supp)
prepost.ggdf$grade <- factor(prepost.ggdf$grade)
levels(prepost.ggdf$grade) <- paste("Grade", levels(prepost.ggdf$grade))
# dev.new()
p2 <- ggplot(prepost.ggdf, aes(x = pre_test, y = post_test)) +
  geom_point(aes(color = supp)) +
  geom_abline(aes(intercept = beta1 + beta2 * (as.numeric(supp)-1),
                  slope = beta3, color = supp)) +
  scale_color_manual(values = c("darkgray", "black")) +
  facet_wrap(~ grade, nrow = 1, scales = "free") +
  scale_x_continuous("pre-test score") +
  scale_y_continuous("post-test score") +
  theme(legend.position = "null")
print(p2)
```

