

Exercise 1 Solutions

Nick Graetz

Sept 16, 2020

Problem 1 (a)

```
# Load data

# Using dplyr:
dataset <- read.delim("C:/Users/ncgra/Downloads/SCHIZREP.dat", header=FALSE, sep="")
dataset <- dataset %>%
  rename(id = V1, imps79 = V2, week = V3, treatment = V4, sex = V5)

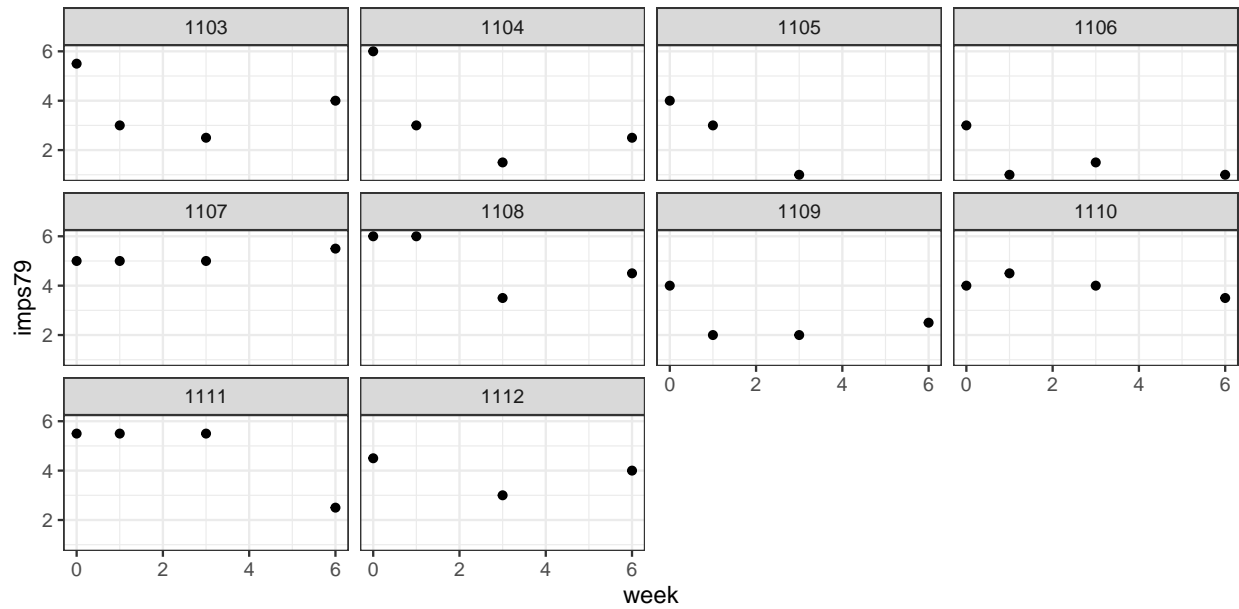
# Using data.table (I use data.table throughout):
dataset <- fread("C:/Users/ncgra/Downloads/SCHIZREP.dat")
setnames(dataset, c('id', 'imps79', 'week', 'treatment', 'sex'))

# Get summary stats over all times
sums <- dataset[,list(imps79_mean=round(mean(imps79),2),
                      imps79_sd=round(sd(imps79),2),
                      N=.N),
                by='week']
sums <- sums[order(week)]
kable(sums)
```

week	imps79_mean	imps79_sd	N
0	5.37	0.87	434
1	4.57	1.23	426
2	4.18	1.83	14
3	4.02	1.44	374
4	3.07	1.70	11
5	3.22	1.72	9
6	3.31	1.48	335

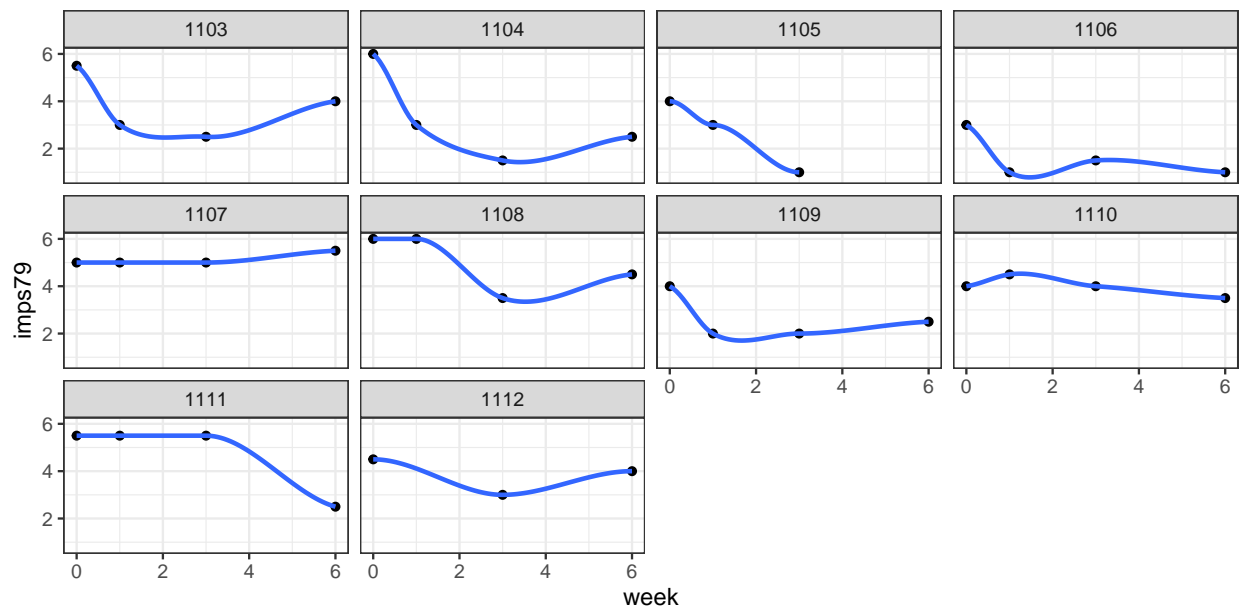
Problem 1 (b)

```
## Grab ids of the first ten patients
first_ten <- unique(dataset[,id])[1:10]
ggplot(data=dataset[id %in% first_ten,]) +
  geom_point(aes(x=week,
                 y=imps79)) +
  facet_wrap(~id) +
  theme_bw()
```



Problem 1 (c)

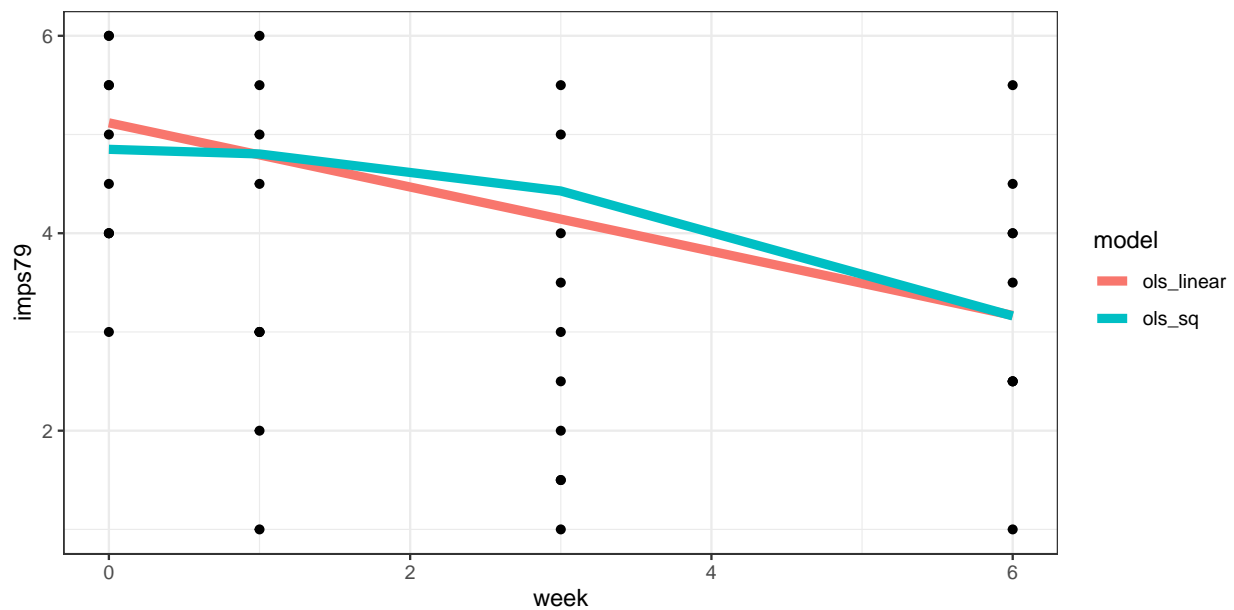
```
ggplot(data=dataset[id %in% first_ten,],
       aes(x=week,
           y=imps79)) +
  geom_point() +
  geom_smooth(method = 'loess') +
  facet_wrap(~id) +
  theme_bw()
```



Problem 1 (c)

```
dataset[, week2 := week^2]
dataset[, ols_linear := predict(lm(imps79~week))]
dataset[, ols_sq := predict(lm(imps79~week2))]
dataset_plot <- melt(dataset,
                      id.vars=c('id','week'),
                      measure.vars=c('ols_linear','ols_sq'),
                      variable.name='model')

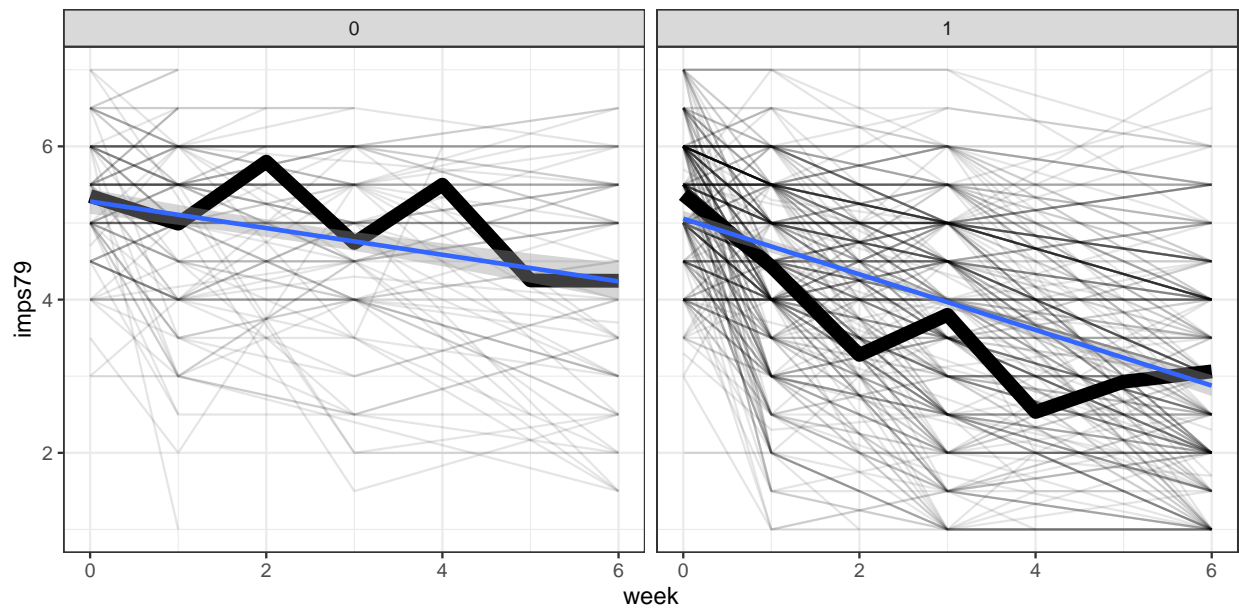
ggplot() +
  geom_point(data=dataset[id %in% first_ten,],
            aes(x=week,
                y=imps79)) +
  geom_line(data=dataset_plot[id %in% first_ten,],
            aes(x=week,
                y=value,
                color=model),
            size=2) +
  theme_bw()
```



Problem 1 (d)

```
sums <- dataset[,list(imps79_mean=mean(imps79)), by=c('week','treatment')]
ggplot(data=dataset,
      aes(x=week,
          y=imps79)) +
  geom_line(aes(group=id),
            alpha=0.1) +
  geom_line(data=sums,
            aes(x=week,
                y=imps79_mean),
            size=3) +
```

```
geom_smooth(method='lm') +
facet_wrap(~treatment) +
theme_bw()
```



Problem 2 (a)

```
library(plm)
library(wooldridge)
data('airfare')
## Our panel is i="id", t="year"
pooled_ols <- lm(log(fare)~as.factor(year)+concen+ldist+ldistsq, data=airfare)
## What is the % increase in fare associated with a 0.1 increase in concen?
pooled_ols_fare_increas <- (exp(pooled_ols$coefficients[['concen']]*0.10)-1)*100
```

The increase expected increase in fare associated with $\Delta\text{concen} = 0.10$ is 3.67%.

Problem 2 (b)

```
confint(pooled_ols)['concen',]
```

```
##      2.5 %      97.5 %
## 0.3011705 0.4190702
```

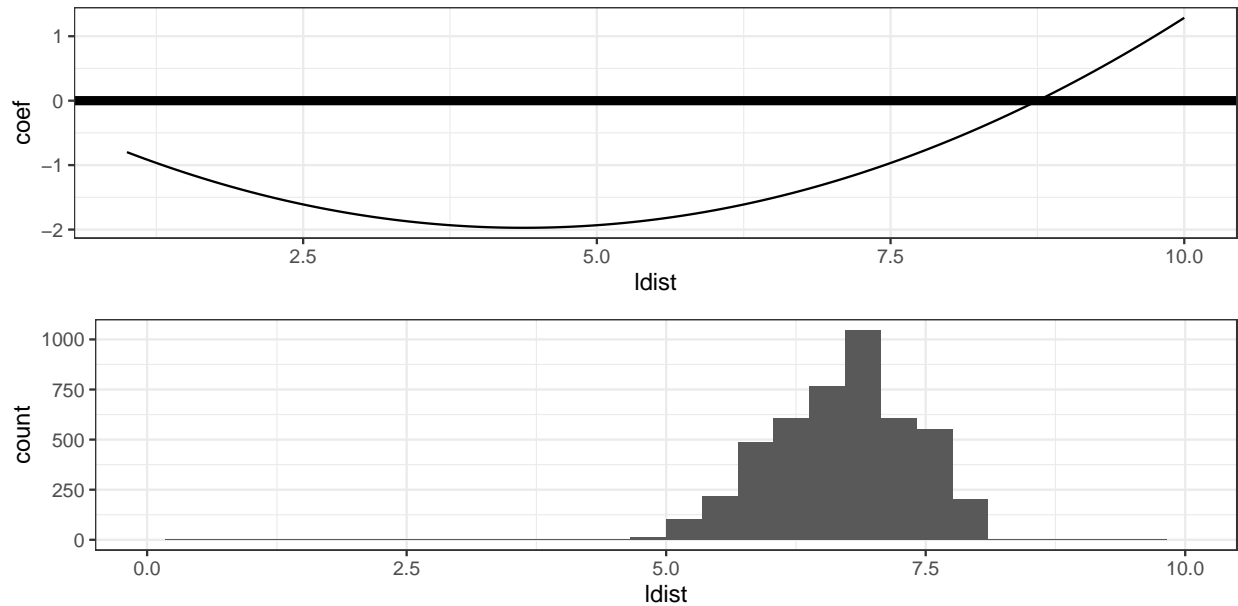
Problem 2 (c)

```
dist_preds <- data.table(ldist=seq(1,10,0.01))
dist_preds[, coef := ldist*pooled_ols$coefficients[['ldist']] +
                  (ldist^2)*pooled_ols$coefficients[['ldistsq']]]
gg1 <- ggplot() +
  geom_line(data=dist_preds,
```

```

    aes(x=ldist,
        y=coef)) +
  geom_hline(yintercept = 0, color='black', size=2) +
  theme_bw()
gg2 <- ggplot() +
  geom_histogram(data=airfare,
                aes(x=ldist)) +
  xlim(c(0,10)) +
  theme_bw()
grid.arrange(grobs=list(gg1,gg2))

```



Problem 2 (d)

```

# Random effects models with plm
res_model <- plm(log(fare)~concen+ldist+ldistsq+as.factor(year),
                data = airfare,
                index = c("id"),
                model = "random")
res_fare_increase <- (exp(res_model$coefficients[['concen']]*0.10)-1)*100
# Try LME instead
library(nlme)
res_model2 <- lme(log(fare)~concen+ldist+ldistsq+as.factor(year),
                 data=airfare, random= ~1|id,
                 method="ML", control=list(opt = "optim"))
res_fare_increase2 <- (exp(res_model2$coefficients$fixed[['concen']]*0.10)-1)*100
# Virtually identical answers

```

The increase expected increase in fare associated with $\Delta\text{concen} = 0.10$ is 2.11%.

Problem 2 (e)

```
# Fixed effects models
library(plm)
fes_model <- plm(log(fare)~concen+ldist+ldistsq+as.factor(year),
                 data = airfare,
                 index = c("id"),
                 model = "within")
fes_fare_increase <- (exp(fes_model$coefficients[['concen']]*0.10)-1)*100
fes_fare_increase

## [1] 1.702927

# Compare to dummy approach
fes_model <- lm(log(fare)~concen+ldist+ldistsq+as.factor(year)+as.factor(id), data=airfare)
fes_fare_increase <- (exp(fes_model$coefficients[['concen']]*0.10)-1)*100
fes_fare_increase

## [1] 1.702927
```

The increase expected increase in fare associated with $\Delta\text{concen} = 0.10$ is 1.7%. As λ approaches 1, FEs and REs estimators will give the same results. Here, λ is 0.835 which is fairly close to 1.

Problem 2 (f)

Possible other time-constant characteristics of routes: populations in origins/destinations, magnitude of total routes at airports, passenger traffic, etc.

Problem 2 (g)

Convinced? It's always possible that we haven't controlled for important time-varying confounders. For this causal question, the FEs model is more conservative than the REs model because it is difficult to check the validity of assumptions of the REs model, such as a lack or correlation between observed variables and omitted variables.