MLM Nested Project D

Xinming Dai, Chongjun Liao, Jeremy Lu, Yu Wang

Compiled on Fri Apr 29 10:32 EDT

Question 1: data generating process

```
set.seed(2042001)

# variance of the random effect
sigma_eta_2 <- 2
sigma_epsilon_2 <- 2

# generate data
dat <-
  tibble(classid = rep(c(1:100), each = 200),
         studentid = 1:(100*200),
         x = runif(100*200, min = 0, max = 1),
         eta_j = rep(rnorm(100, sd = sqrt(sigma_eta_2)), each = 200),
         epsilon = rnorm(100*200, sd = sqrt(sigma_epsilon_2)),
         y = x + eta_j + epsilon)
```

Question 2: fit the model

```
lmer_fit1 <- lmer(y ~ x + (1|classid), data = dat)
summary(lmer_fit1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y ~ x + (1 | classid)
##    Data: dat
##
## REML criterion at convergence: 71227.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0143 -0.6761  0.0024  0.6711  3.7584
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  classid  (Intercept) 1.893    1.376
##  Residual             2.008    1.417
```

```
## Number of obs: 20000, groups:  classid, 100
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept) -7.493e-03  1.391e-01  1.022e+02  -0.054    0.957
## x            9.864e-01  3.496e-02  1.990e+04  28.216   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##   (Intr)
## x -0.126
```

Question 2:

a. The estimated coefficient of X is 0.986.
b. The 95% confidence interval for this coefficient estimate is $[0.986 - 1.96 * 0.035, 0.986 + 1.96 * 0.035] =$ [0.9174, 1.0546] cover the true coefficient, which is 1.

Question 3:

Question 4:

```
# 4a
z <- rbinom(100*200,1,dat$x)
# 4b
dat$y_q4 <- ifelse(z==1,NA,dat$y)
# 4c
lmer_fit_mar <- lmer(y_q4 ~ x + (1|classid), data = dat)
summary(lmer_fit_mar)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y_q4 ~ x + (1 | classid)
##    Data: dat
##
## REML criterion at convergence: 35559
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9126 -0.6748  0.0037  0.6705  3.7484
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  classid  (Intercept) 1.853    1.361
##  Residual             2.004    1.415
## Number of obs: 9936, groups:  classid, 100
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1.299e-03  1.384e-01 1.035e+02   0.009    0.993
```

```
## x           9.764e-01  6.065e-02 9.838e+03  16.100   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##   (Intr)
## x -0.148
```

d.

i. The 95% confidence interval is [0.858,1.095], which covers the "truth".

e.

```
N <- sum(z==0)
```

We use N = 9936 samples in the model fit.

Question 5: