Homework 6

Psych 5068

Emorie Beck

April 14, 2018

Contents

For this assignment, you will extend our analyses of the Curran reading data (Curran_New_Trimmed.csv).

Workspace

Packages

```
library(psych)
library(lme4)
library(knitr)
library(qqplotr)
library(influence.ME)
library(HLMdiag)
library(multcomp)
library(kableExtra)
```

```r
library(plyr)
library(tidyverse)
```

Data

```r
data_url <- "https://raw.githubusercontent.com/emoriebeck/homeworks/master/homework6/Curran_New_Trimmed
dat      <- data_url %>% read.csv %>% tbl_df
```

Question 1

Create dummy codes for each of the four time periods. Call these new variables TD1, TD2, TD3, and TD4.

```r
dat <- dat %>%
  mutate(
    TD1 = ifelse(time == 1, 1, 0),
    TD2 = ifelse(time == 2, 1, 0),
    TD3 = ifelse(time == 3, 1, 0),
    TD4 = ifelse(time == 4, 1, 0)
  )
```

Question 2

Fit a no-intercept model:

Level 1: $read_{ti} = \pi_{0i}TD1_{ti} + \pi_{1i}TD2_{ti} + \pi_{2i}TD3_{ti} + \pi_{3i}TD4_{ti} + e_{ij}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10}$
$\pi_{2i} = \beta_{20}$
$\pi_{3i} = \beta_{30}$

```r
source("https://raw.githubusercontent.com/emoriebeck/homeworks/master/table_fun.R")
fit2 <- lmer(read ~ -1 + TD1 + TD2 + TD3 + TD4 + (-1 + TD1 | id), dat = dat)
tab2 <- table_fun(fit2)

options(knitr.kable.NA = '')
tab2 %>% select(-type) %>%
  kable(., "latex", escape = F, digits = 2, booktabs = T,
        col.names = c("", c("b", "CI"))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,4) %>%
  group_rows("Fit", 6, 7) %>%
  group_rows("Random", 5, 5) %>%
  add_header_above(c(" " = 1, "Model 2" = 2))
```

	Model 2	
	b	CI
Fixed		
TD1	2.55	[2.44, 2.67]
TD2	4.09	[4.02, 4.16]
TD3	5.09	[5.05, 5.30]
TD4	5.80	[5.66, 5.88]
Random		
τ_{00}	0.00	[0.00, 0.37]
R^2_m	0.55	
R^2_c	0.55	

Question 3

Perform follow-up tests on the model.

Part A

Create a matrix of contrasts that will test the linear, quadratic, and cubic trends in the data.

```
round(contr <- t(contr.poly(4)),2)
```

```
##      [,1]  [,2]  [,3] [,4]
## .L -0.67 -0.22  0.22 0.67
## .Q  0.50 -0.50 -0.50 0.50
## .C -0.22  0.67 -0.67 0.22
```

Part B

Test this matrix using `glht( )` in the multcomp package. Which trends are significant?

```
ghlt_fit2   <- multcomp::glht(fit2, linfct = contr)
res_fit2    <- confint(ghlt_fit2, calpha = univariate_calpha())
res_fit2_df <- res_fit2$confint %>% data.frame() %>%
  mutate(term = rownames(.)) %>%
  rename(b = Estimate)

res_fit2_df  %>%
  mutate_at(vars(-term), funs(sprintf("%.2f", .))) %>%
  mutate(term = mapvalues(term, c(".L", ".Q", ".C"),
      c("age\\_10", "age\\_10\\_2", "age\\_10\\_3")),
      CI = sprintf("[%s, %s]", lwr, upr)) %>%
  select(term, b, CI) %>%
  kable(., "latex", booktabs = T, escape = F,
        caption = "Q3b ghlt()", col.names = c("Term", "b", "CI"),
        align = c("l", "c", "c"))
```

There is a linear (b_{linear} 2.4, CI = [2.27, 2.53]) and quadratic ($b_{quadratic}$ -0.41, CI = [-0.54, -0.28]) but not cubic (b_{cubic} 0.06, CI = [-0.07, 0.18]). In other words, reading scores linearly increase with age, but also show a quadratic trend, such that rates of improvement slow over time.

Table 1: Q3b ghlt()

Term	b	CI
age_10	2.40	[2.27, 2.53]
age_10_2	-0.41	[-0.54, -0.28]
age_10_3	0.06	[-0.07, 0.18]

Question 4

Test each of the following models (use ML not REML):

Part A

Model 1:
Level 1: $read_{ti} = \pi_{0i} + e_{ti}$ Level 2: $\pi_{0i} = \beta_{00} + r_{0i}$

```
fit4a <- lmer(read ~ 1 + (1 | id), data = dat)
tab4a <- table_fun(fit4a)
```

Part B

Model 2:
Level 1:
$read_{ti} = \pi_{0i} + \pi_{1i}(age - 10)_{ti} + e_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$

```
dat   <- dat %>% mutate(age_10 = kidagetv - 10)
fit4b <- lmer(read ~  age_10 + (age_10 | id), data = dat)
tab4b <- table_fun(fit4b)
```

Part C

Model 3:
Level 1:
$read_{ti} = \pi_{0i} + \pi_{1i}(age - 10)_{ti} + \pi_{2i}(age - 10)^2_{ti} + e_{ti}$

Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$
$\pi_{2i} = \beta_{20} + r_{2i}$

```
dat   <- dat %>% mutate(age_10_2 = age_10^2)
fit4c <- lmer(read ~  age_10 + age_10_2 + (age_10 + age_10_2 | id), data = dat)
tab4c <- table_fun(fit4c)
```

Part D

Model 4:
Level 1:
$read_{ti} = \pi_{0i} + \pi_{1i}(age - 10)_{ti} + \pi_{2i}(age - 10)^2_{ti} + \pi_{3i}(age - 10)^3_{ti} + e_{ti}$

Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$
$\pi_{2i} = \beta_{20} + r_{2i}$
$\pi_{3i} = \beta_{30} + r_{3i}$

```
dat <- dat %>% mutate(age_10_3 = age_10^3)
fit4d <- lmer(read ~  age_10 + age_10_2 + age_10_3 + (age_10 + age_10_2 + age_10_2 | id), data = dat)
tab4d <- table_fun(fit4d)
```

```
tab4a %>% mutate(model = "4a") %>%
  full_join(tab4b %>% mutate(model = "4b")) %>%
  full_join(tab4c %>% mutate(model = "4c")) %>%
  full_join(tab4d %>% mutate(model = "4d")) %>%
  # mutate(est = ifelse(type != "Fit",
  #   sprintf("\\makecell{%s \\\\ {%s}}", b, CI), b)) %>%
  # select(-b, -CI) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, model, est, sep = ".") %>%
  spread(key = tmp, value = value) %>%
  select(-type) %>%
  kable(., "latex", escape = F, digits = 2, booktabs = T,
        col.names = c("", rep(c("b", "CI"), times = 4))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,4) %>%
  group_rows("Fit", 5, 6) %>%
  group_rows("Random", 7, 9) %>%
  add_header_above(c(" " = 1, "Model 4a" = 2, "Model 4b" = 2,
                     "Model 4c" = 2, "Model 4b" = 2))
```

	Model 4a		Model 4b		Model 4c		Model 4b	
	b	CI	b	CI	b	CI	b	CI
Fixed								
age_10			0.56	[0.55, 0.58]	0.53	[0.52, 0.56]	0.50	[0.46, 0.53]
age_10_2					-0.05	[-0.05, -0.05]	-0.05	[-0.05, -0.04]
age_10_3							0.00	[0.00, 0.01]
Intercept	4.15	[4.08, 4.19]	4.43	[4.39, 4.53]	4.67	[4.63, 4.75]	4.66	[4.63, 4.75]
R^2_c	0.11		0.87		0.91		0.91	
R^2_m	0.00		0.58		0.60		0.60	
Random								
τ_{00}	0.29	[0.13, 0.43]	0.85	[0.70, 1.02]	1.12	[1.00, 1.23]	1.12	[0.96, 1.30]
τ_{11}			0.02	[0.01, 0.02]	0.02	[0.01, 0.02]	0.02	[0.01, 0.02]
τ_{22}					0.00	[0.00, 0.00]	0.00	[0.00, 0.00]

Part E

Using likelihood ratio tests, compare the fit of these models. Which fits the data the best?

```r
anova(fit4a, fit4b, fit4c, fit4d)
```

```
## Data: dat
## Models:
## fit4a: read ~ 1 + (1 | id)
## fit4b: read ~ age_10 + (age_10 | id)
## fit4c: read ~ age_10 + age_10_2 + (age_10 + age_10_2 | id)
## fit4d: read ~ age_10 + age_10_2 + age_10_3 + (age_10 + age_10_2 + age_10_2 |
## fit4d:     id)
##       Df    AIC    BIC  logLik deviance     Chisq Chi Df Pr(>Chisq)
## fit4a  3 4428.1 4443.3 -2211.1   4422.1
## fit4b  6 2855.3 2885.6 -1421.6   2843.3 1578.8036      3    < 2e-16 ***
## fit4c 10 2641.1 2691.7 -1310.6   2621.1  222.1635      4    < 2e-16 ***
## fit4d 11 2636.8 2692.4 -1307.4   2614.8    6.3072      1    0.01203 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Baased on the likelihood ratio tests, model 4 (with the linear, quadratic, and cubic terms) fit the data best.

Question 5

Compare what you found in the two approaches (discrete time and continuous age).

```r
res_fit2_df %>%
  mutate_at(vars(-term), funs(sprintf("%.2f", .))) %>%
  mutate(term = mapvalues(term, c(".L", ".Q", ".C"),
      c("age\\_10", "age\\_10\\_2", "age\\_10\\_3")),
      CI = sprintf("[%s, %s]", lwr, upr),
      type = "Fixed Parts", Model = "Q2") %>%
  select(type, term, b, CI, Model) %>%
  full_join(tab4d %>% filter(type == "Fixed Parts") %>%
            mutate(Model = "Q4")) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, Model, est, sep = ".") %>%
  spread(key = tmp, value = value) %>%
  select(-type) %>%
  kable(., "latex", escape = F, booktabs = T) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,4) %>%
  add_header_above(c(" " = 1, "Discrete" = 2, "Continous" = 2))
```

	Discrete		Continous	
term	Q2.b	Q2.CI	Q4.b	Q4.CI
Fixed				
age_10	2.40	[2.27, 2.53]	0.50	[0.46, 0.53]
age_10_2	-0.41	[-0.54, -0.28]	-0.05	[-0.05, -0.04]
age_10_3	0.06	[-0.07, 0.18]	0.00	[0.00, 0.01]
Intercept			4.66	[4.63, 4.75]

Part A

Are there any important differences in the inferences you would draw? The magnitude of the differences of the slopes are quite different. In the discrete time model, there appears to be a quadratic but not a cubic trend in the data, while the continuous age model has cubic trends. The magnitude of the change is also quite different. The linear age term of the discrete time model is nearly 5 times that of the continuous age model.

Part B

What might account for the different findings? The scale of timing as well as the individuals included at each time point are different across models. Ages range from 6 to 14 (8 years), while time ranges across 4 waves. Moreover, age was centered at 10, while time was not centered, so the intercepts of the moel will change.

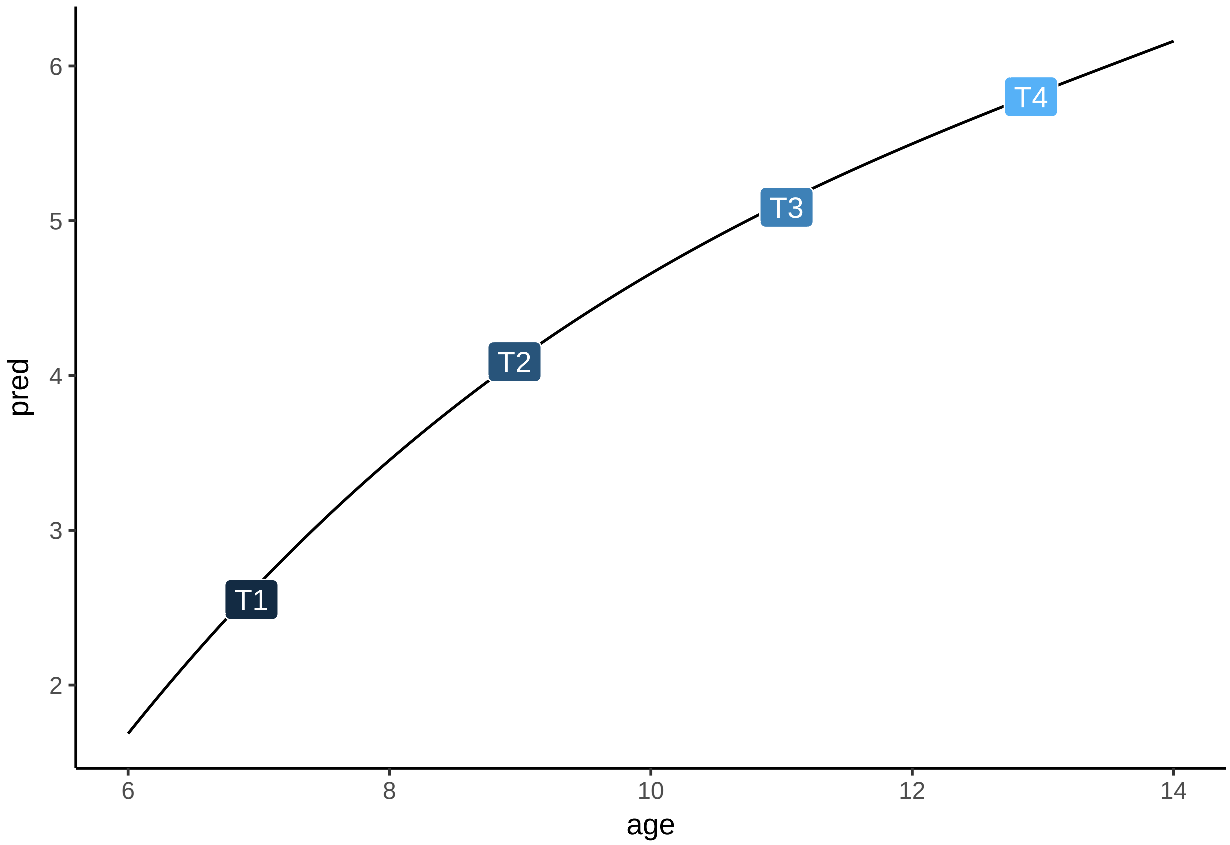
Part C

Plot the results of Model 4. Use Age on the abscissa, ranging from 6 to 14. On the same figure, add the mean reading scores for TD1, TD2, TD3, and TD4 (the intercepts from Question 2). Locate these means along the abscissa according to the average age of the kids at each time period. Does this plot offer any further insights relevant to (b)?

```r
pred_dat <- tibble(
  age = seq(6, 14, .01),
  age_10 = age - 10,
  age_10_2 = age_10^2,
  age_10_3 = age_10^3
  ) %>%
  mutate(pred = predict(fit4d, newdata = ., re.form = NA))

mean_dat <- dat %>%
  group_by(time) %>%
  summarize_at(vars(read, kidagetv), funs(mean(., na.rm = T)))

pred_dat %>%
  ggplot(aes(x = age, y = pred)) +
    geom_line() +
    geom_label(data = mean_dat, color = "white",
               aes(x = kidagetv, y = read, label = paste0('T',time), fill = time)) +
    theme_classic() +
    theme(legend.position = "none")
```



The plot suggests that age increased over time, such that the linear and quadratic trends evident in both models reflect changes in age of the sample. But as noted previously, the different scale (1,2,3,4) will influence the size of the slopes and the intercepts.

Question 6

Using Model 3, add the age of the mother (momage) as a moderator:
Level 1: $read_{ti} = \pi_{0i} + \pi_{1i}(age - 10)_{ti} + \pi_{2i}(age - 10)^2_{ti} + e_{ti}$
Level 2: $\pi_{0i} = \beta_{00} + \beta_{01}momage_i + r_{0i}$
$\pi_{1i} = \beta_{10} + \beta_{11}momage_i + r_{1i}$
$\pi_{2i} = \beta_{20} + \beta_{21}momage_i + r_{2i}$

```
fit6  <- lmer(read ~  age_10*momage + age_10_2*momage + (age_10 + age_10_2 | id), data = dat)
tab_6 <- table_fun(fit6)

tab_6 %>% select(-type) %>%
  kable(., "latex", escape = F, digits = 2, booktabs = T,
        col.names = c("", c("b", "CI"))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fit", 10, 11) %>%
  group_rows("Fixed", 1,6) %>%
  group_rows("Random", 7, 9) %>%
  add_header_above(c(" " = 1, "Model 6" = 2))
```

	Model 6	
	b	CI
Fixed		
Intercept	1.41	[-0.36, 3.13]
age_10	0.03	[-0.27, 0.43]
momage	0.13	[0.06, 0.19]
age_10_2	0.01	[-0.07, 0.11]
age_10:momage	0.02	[0.00, 0.03]
momage:age_10_2	-0.00	[-0.01, 0.00]
Random		
τ_{00}	1.07	[1.05, 1.37]
τ_{11}	0.02	[0.01, 0.02]
τ_{22}	0.00	[0.00, 0.00]
R^2_m	0.61	
R^2_c	0.91	

Part A

Are any of the interactions involving momage significant? If so, explain what they mean. The interation between the linear age term and age of the mother is significant. In other words, reading ability varies as a function of both the age of the child and the age of the mother at baseline. Students who are farther above the age 10 with older mothers show steeper increases in reading age than those below age 10.

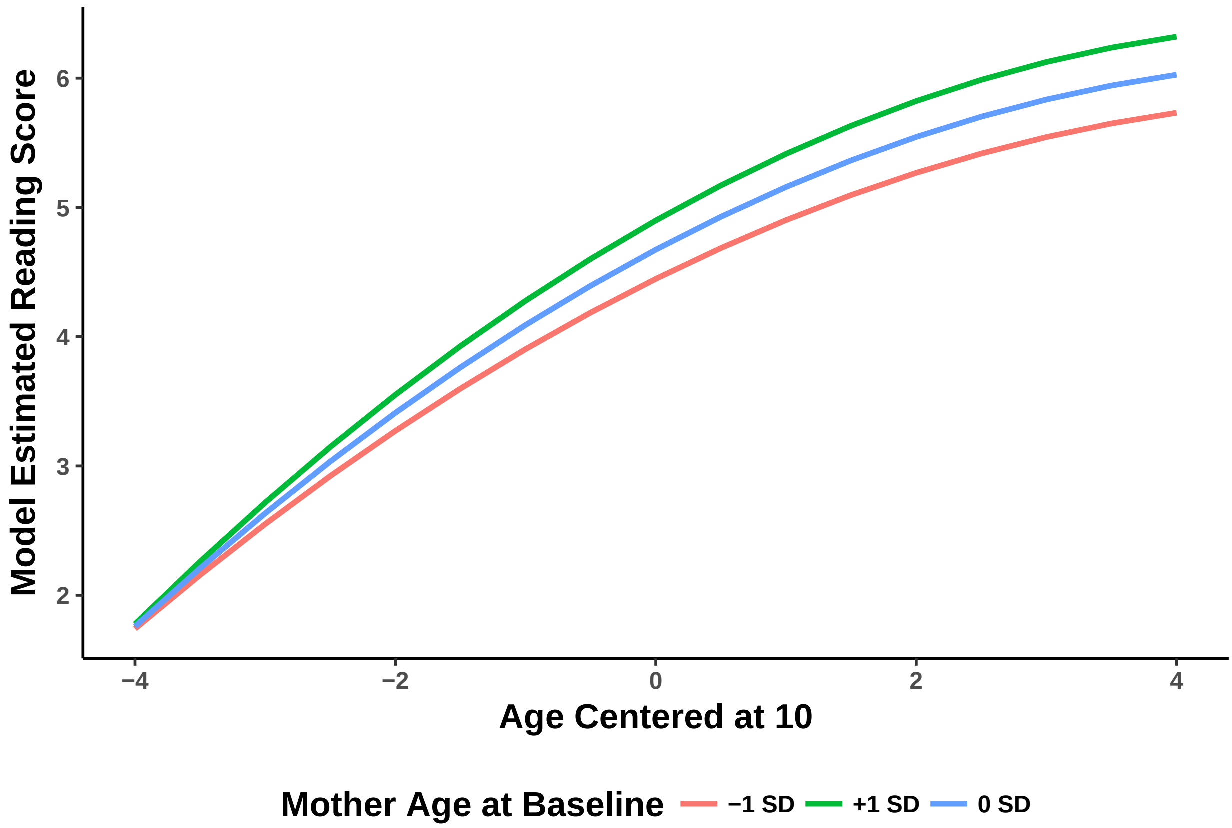
Part B

Illustrate this new model by plotting the age-reading relationship separately for mothers 1 standard deviation below the mean for momage and 1 standard deviation above the mean for momage.

```
means <- dat %>% summarize_at(vars(momage),
            funs(mean = mean(., na.rm = T),
                 sd = sd(., na.rm = T)))

crossing(
  age_10 = seq(-4,4,.5),
  momage = c(means$mean, means$mean - means$sd, means$mean + means$sd)
) %>% mutate(age_10_2 = age_10^2) %>%
  mutate(pred = predict(fit6, newdata = ., re.form = NA),
         momage = mapvalues(momage, unique(momage), c("-1 SD", "0 SD", "+1 SD"))) %>%

  ggplot(aes(x = age_10, y = pred, color = momage)) +
    geom_line(size = 1) +
    labs(x = "Age Centered at 10", y = "Model Estimated Reading Score", color = "Mother Age at Baseline
    theme_classic() +
    theme(legend.position = "bottom",
          axis.text = element_text(face = "bold"),
          axis.title = element_text(face = "bold", size = rel(1.2)),
          legend.text = element_text(face = "bold"),
          legend.title = element_text(face = "bold", size = rel(1.2)))
```


Model Estimated Reading Score
6
5
4
3
2
−4
−2
0
2
4
Age Centered at 10
Mother Age at Baseline
−1 SD
+1 SD
0 SD