Homework 7

Psych 5068

Emorie Beck

March 26, 2018

Contents

Workspace **1**
Packages . . . 1
Data . . . 1

Question 1 **2**

Question 2 **2**
Part A . . . 3
Part B . . . 3

Question 3 **4**
Part A . . . 5
Part B . . . 5

Question 4 **6**
Part A . . . 7
Part B . . . 7

Question 5 **8**
Part A . . . 9
Part B . . . 9

Question 6 **9**
Part A . . . 10
Part B . . . 10

Question 7 **11**
Part A . . . 11
Part B . . . 11
Part C . . . 11

Workspace

Packages

Data

```
source("https://raw.githubusercontent.com/emoriebeck/homeworks/master/table_fun.R")
data_url <- "https://raw.githubusercontent.com/emoriebeck/homeworks/master/homework7/daily_diary(4).csv
dat      <- data_url %>% read.csv %>% tbl_df %>% select(-X)
```

Question 1

1. Begin by creating lagged variables for self-esteem (rse lag 1), the reported average positive events (posmn lag 1), and the reported average negative events (negmn lag 1). You may find functions that can help you do this or you may decide to do this with your own code. Either way, you will need to do this carefully because the reports by participants do not always reflect consecutive days. That is, a person might provide data on the following days: 1, 2, 3, 6, 8, 9, 10, 12, 13, and 14. Lagged variables would only be available for days 2, 3, 9, 10, 13, and 14. Those are the only days when there was also a measure available on the immediately preceding day. So, we will lose some data in the process of creating lagged variables but that will insure that the meaning of the lagged variable is consistent.

```
dat <- dat %>%
  mutate(cesd_z = as.numeric(scale(cesd))) %>%
  full_join(crossing(id = unique(dat$id), day = seq(1, max(dat$id)))) %>%
  arrange(id, day) %>%
  group_by(id) %>%
  mutate_at(vars(rse, posmn, negmn), funs(lag_1 = lag(., n = 1)))
```

Question 2

First, let's see how persistent self-esteem is for all of the participants.

Level 1:
$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$

```
fit2     <- lmer(rse ~ 1 + rse_lag_1 + (rse_lag_1 | id), data = 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", 1,2) %>%
  group_rows("Fit", 5, 6) %>%
  group_rows("Random", 3, 4) %>%
  add_header_above(c(" " = 1, "Model Q2" = 2))
```

	Model Q2	
	b	CI
Fixed		
Intercept	3.56	[3.27, 3.70]
rse_lag_1	0.34	[0.31, 0.39]
Random		
τ_{00}	0.77	[0.53, 1.20]
τ_{11}	0.01	[0.00, 0.02]
R^2_m	0.15	
R^2_c	0.48	

Part A

What is the value of the fixed effect for rse lag 1? Is it significant?
The fixed effect for rse_lag_1 is significant, 0.34, 95% CI [0.31, 0.39].

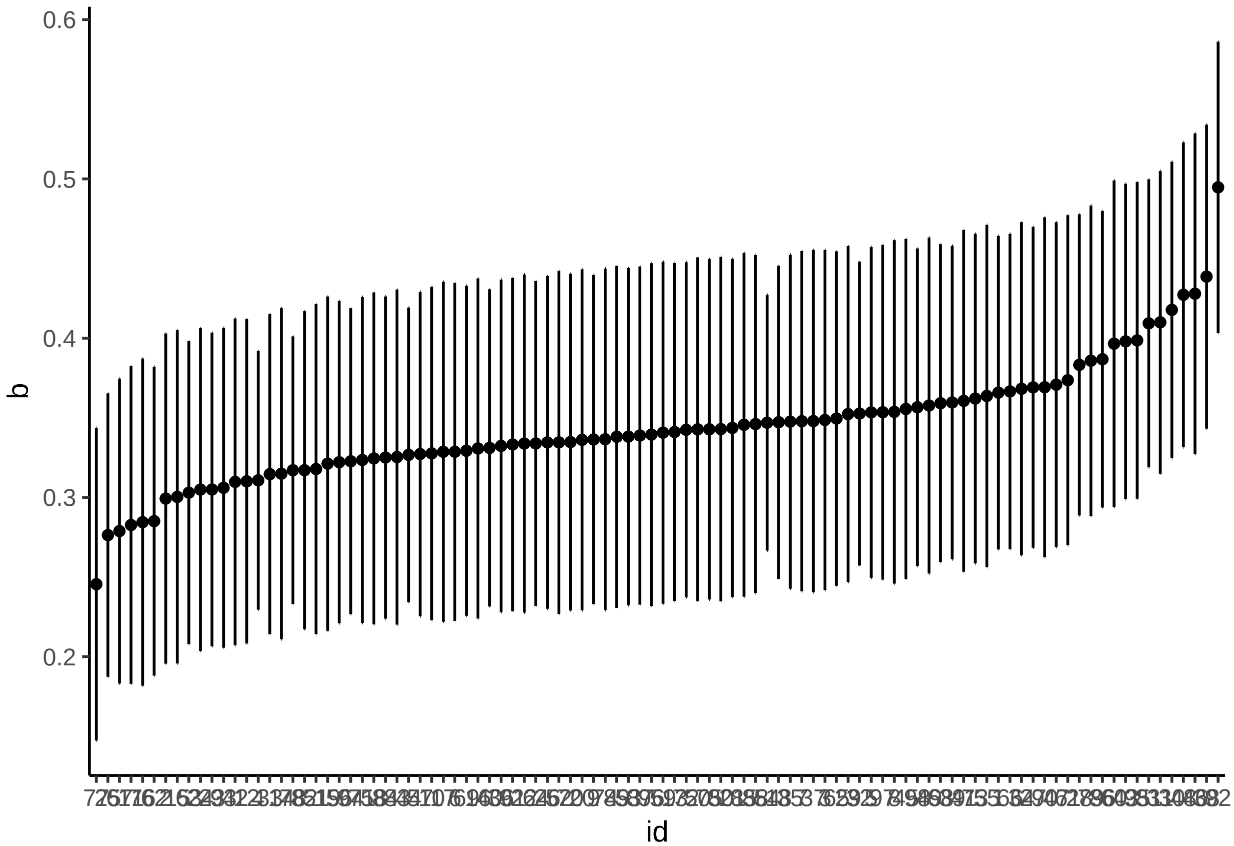
Part B

Provide a caterpillar plot showing the individual estimated coefficients for rse lag 1. The individual coefficients can be obtained using the coef( ) function for the fit object. Each individual coefficient has a standard error that can be obtained with the se.coef( ) function, a helper function from the arm package. Is persistence in self-esteem a common occurrence? What makes you say so?

```r
coefs2 <- cbind(coef(fit2)[["id"]], arm::se.coef(fit2)[["id"]]) %>% data.frame %>%
  setNames(c("Intercept.b", "rse_lag_1.b", "Intercept.se", "rse_lag_1.se")) %>%
  mutate(id = rownames(.)) %>%
  tbl_df %>%
  gather(key = term, value = value, -id) %>%
  separate(term, c("term", "tmp"), sep = "[.]") %>%
  spread(key = tmp, value = value) %>%
  filter(term == "rse_lag_1")

orders <- coefs2 %>% arrange(b)

coefs2 %>% mutate(id = factor(id, levels = orders$id)) %>%
  ggplot(aes(x = id, y = b)) +
    geom_point() +
    geom_errorbar(aes(ymin = b - se, ymax = b + se), width = 0) +
    # geom_hline(aes(yintercept = fixef(fit3)["posmn_lag_1"]), color = "red", size = 1) +
    theme_classic()
```



```
perc_above <- coefs2 %>% mutate(lower = b - 1.96*se, sign = sign(lower)) %>%
  group_by(sign) %>% summarize(n = n()) %>% ungroup() %>% mutate(n = n/sum(n))
```

Yes, across, the entire, sample, the lagged effect is positive. In other words, self-esteem at time t (the lagged variable) is associated with higher self-esteem at time t + 1.

Question 3

Now let's see how persistent positive events are for all of the participants.

Level 1:
$posmn_{ti} = \pi_{0i} + \pi_{1i}posmn\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$

```
fit3     <- lmer(posmn ~ 1 + posmn_lag_1 + (posmn_lag_1 | id), data = dat)
tab3 <- table_fun(fit3)

options(knitr.kable.NA = '')
tab3 %>% select(-type) %>%
  kable(., "latex", escape = F, digits = 2, booktabs = T,
        col.names = c("", c("b", "CI"))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,2) %>%
  group_rows("Fit", 5, 6) %>%
```

```r
  group_rows("Random", 3, 4) %>%
  add_header_above(c(" " = 1, "Model Q2" = 2))
```

	Model Q2	
	b	CI
Fixed		
Intercept	0.89	[0.69, 0.96]
posmn_lag_1	0.25	[0.21, 0.31]
Random		
τ_{00}	0.13	[0.10, 0.23]
τ_{11}	0.03	[0.00, 0.08]
R^2_m	0.08	
R^2_c	0.50	

Part A

Is there significant persistence for positive events? How would you interpret this effect? Speculate about why it might occur.

The fixed effect for posmn_lag_1 is significant, 0.25, 95% CI [0.21, 0.31]. In other words, positive events at time t + 1 (the lagged variable) predicts positive events at the previous time point. This may occur becuase of environmental or psychological carryover effects, which we often term selection effects. Someone who experiences positive events may live in an environment where positive events are more likely. Moreover, someone who experiences positive events may select into future positive events in order to maintain positive affect.

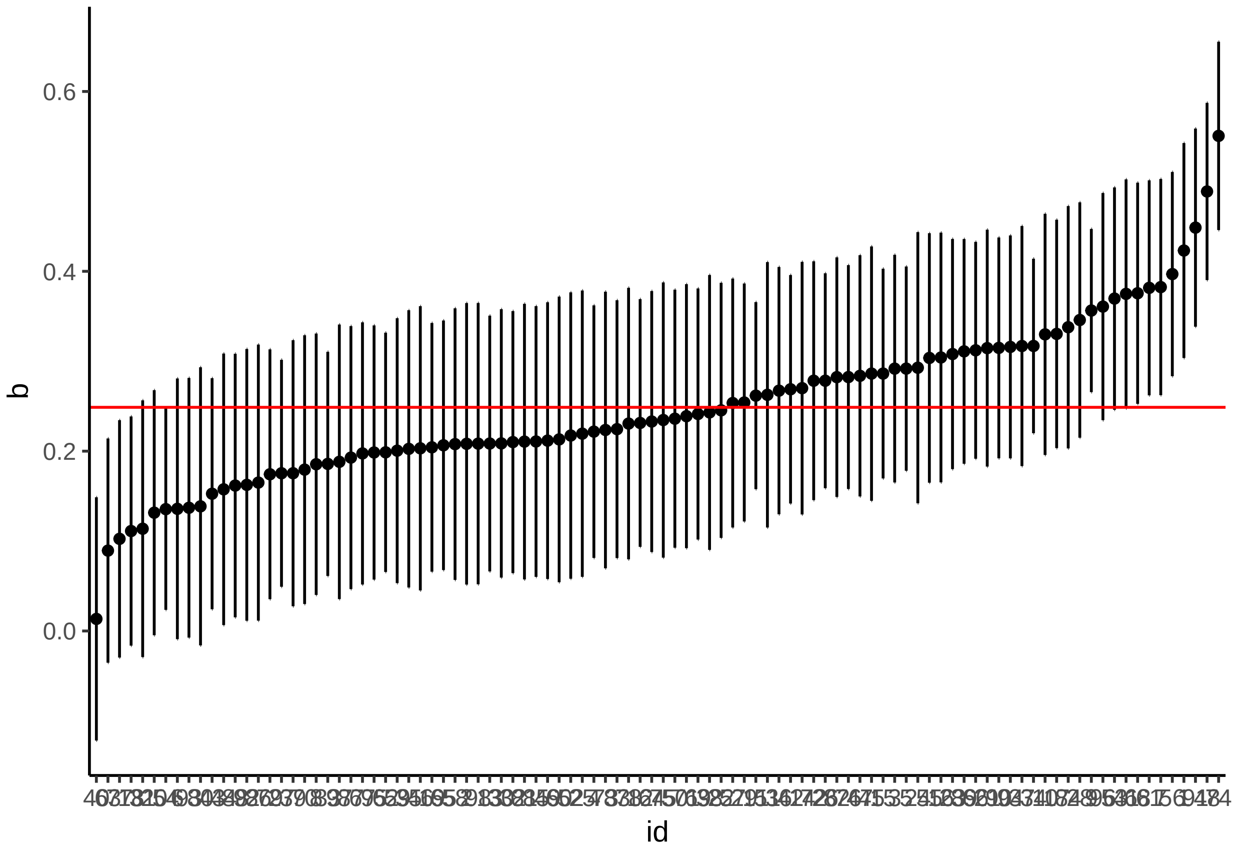
Part B

Provide a caterpillar plot of the individual estimated coefficients for posmn lag 1. Is this form of persistence common?

```r
coefs3 <- cbind(coef(fit3)[["id"]], arm::se.coef(fit3)[["id"]]) %>% data.frame %>%
  setNames(c("Intercept.b", "posmn_lag_1.b", "Intercept.se", "posmn_lag_1.se")) %>%
  mutate(id = rownames(.)) %>%
  tbl_df %>%
  gather(key = term, value = value, -id) %>%
  separate(term, c("term", "tmp"), sep = "[.]") %>%
  spread(key = tmp, value = value) %>%
  filter(term == "posmn_lag_1")

orders <- coefs3 %>% arrange(b)

coefs3 %>% mutate(id = factor(id, levels = orders$id)) %>%
  ggplot(aes(x = id, y = b)) +
    geom_point() +
    geom_errorbar(aes(ymin = b - se, ymax = b + se), width = 0) +
    geom_hline(aes(yintercept = fixef(fit3)["posmn_lag_1"]), color = "red") +
    theme_classic()
```




```
perc_above <- coefs3 %>% mutate(lower = b - 1.96*se, sign = sign(lower)) %>%
  group_by(sign) %>% summarize(n = n()) %>% ungroup() %>% mutate(n = n/sum(n))
```

Yes, this persistence is somewhat common but not greatly reliable. The lower bound of the confidence intervals on individual level slopes exceeds zero for 38% of people.

Question 4

Finally, let's see how persistent negative events are for all of the participants.

Level 1:
$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$

```
fit4     <- lmer(negmn ~ 1 + negmn_lag_1 + (negmn_lag_1 | id), data = dat)
tab4 <- table_fun(fit4)

options(knitr.kable.NA = '')
tab4 %>% select(-type) %>%
  kable(., "latex", escape = F, digits = 2, booktabs = T,
        col.names = c("", c("b", "CI"))) %>%
  kable_styling(full_width = F) %>%
  group_rows("Fixed", 1,2) %>%
  group_rows("Fit", 5, 6) %>%
```

```r
  group_rows("Random", 3, 4) %>%
  add_header_above(c(" " = 1, "Model Q4" = 2))
```

	Model Q4	
	b	CI
Fixed		
Intercept	0.36	[0.31, 0.39]
negmn_lag_1	0.32	[0.28, 0.39]
Random		
τ_{00}	0.05	[0.03, 0.07]
τ_{11}	0.03	[0.02, 0.04]
R^2_m	0.13	
R^2_c	0.45	

Part A

Is there significant persistence for negative events?
The fixed effect for negmn_lag_1 is significant, , 95% CI .

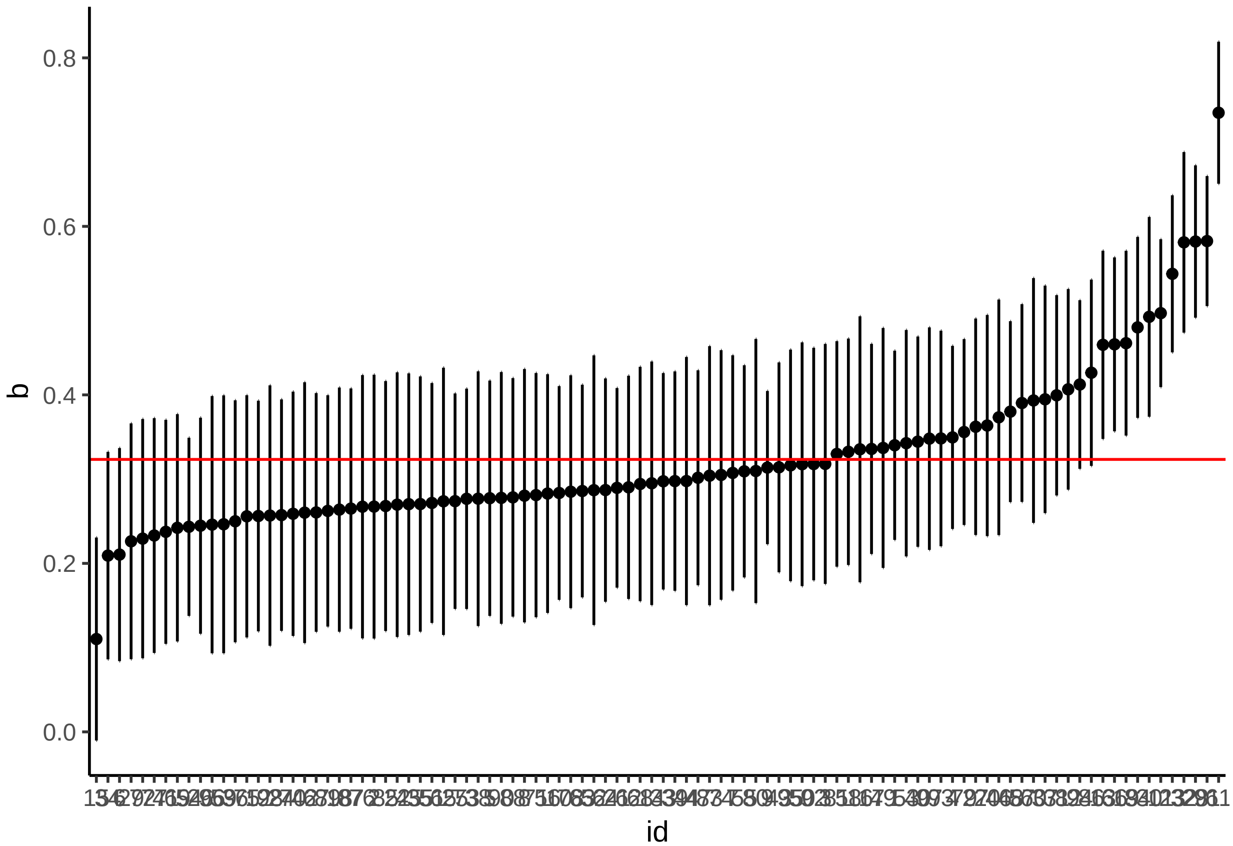
Part B

Provide a caterpillar plot of the individual estimated coefficients for negmn lag 1 and comment on how common this form of persistence is among participants.

```r
coefs4 <- cbind(coef(fit4)[["id"]], arm::se.coef(fit4)[["id"]]) %>% data.frame %>%
  setNames(c("Intercept.b", "negmn_lag_1.b", "Intercept.se", "negmn_lag_1.se")) %>%
  mutate(id = rownames(.)) %>%
  tbl_df %>%
  gather(key = term, value = value, -id) %>%
  separate(term, c("term", "tmp"), sep = "[.]") %>%
  spread(key = tmp, value = value) %>%
  filter(term == "negmn_lag_1")

orders <- coefs4 %>% arrange(b)

coefs4 %>% mutate(id = factor(id, levels = orders$id)) %>%
  ggplot(aes(x = id, y = b)) +
    geom_point() +
    geom_errorbar(aes(ymin = b - se, ymax = b + se), width = 0) +
    geom_hline(aes(yintercept = fixef(fit4)["negmn_lag_1"]), color = "red") +
    theme_classic()
```




```
perc_above <- coefs4 %>% mutate(lower = b - 1.96*se, sign = sign(lower)) %>%
  group_by(sign) %>% summarize(n = n()) %>% ungroup() %>% mutate(n = n/sum(n))
```

Yes, this persistence is common. The lower bound of the confidence intervals on individual level slopes exceeds zero for 64% of people.

Question 5

Now fit a model with self-esteem as the outcome and with lagged self-esteem, current positive events, and current negative events as predictors:

Level 1:
$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn\_lag\_1_{ti} + \pi_{3i} negmn\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$
$\pi_{2i} = \beta_{20} + r_{2i}$
$\pi_{3i} = \beta_{30} + r_{3i}$

```
fit5 <- lmer(rse ~ rse_lag_1 + posmn_lag_1 + negmn_lag_1 +
    (rse_lag_1 + posmn_lag_1 + negmn_lag_1 | id), data = dat)

tab5 <- table_fun(fit5)

options(knitr.kable.NA = '')
tab5 %>% 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", 9, 10) %>%
  group_rows("Random", 5, 8) %>%
  add_header_above(c(" " = 1, "Model Q5" = 2))
```

	Model Q5	
	b	CI
Fixed		
Intercept	3.93	[3.66, 4.21]
rse_lag_1	0.30	[0.27, 0.33]
posmn_lag_1	-0.02	[-0.06, 0.09]
negmn_lag_1	-0.19	[-0.33, -0.03]
Random		
τ_{00}	0.50	[0.22, 1.16]
τ_{11}	0.00	[0.00, 0.01]
τ_{22}	0.02	[0.00, 0.10]
τ_{33}	0.02	[0.00, 0.07]
R_m^2	0.15	
R_c^2	0.49	

Part A

This model provides a different way of examining the impact of current events on self- esteem. By partialling the previous day's self-esteem, how would the influence of current events be interpreted in this model?

By partialling out the previous day's self esteem, we are able to look at the unique lagged effect of negative and positive emotion on the self-esteem.

Part B

Fit the model and comment on the magnitude and interpretation of the coefficients for positive and negative events.

There is a negative effect of residual negative events on self-esteem (b = ; 95% CI). The previous day's negative events predict lower self esteem.

There is no effect of residual positive events on self-esteem (b = ; 95% CI). The previous day's positive events do not predict lower self esteem.

Question 6

Fit a full lagged model.

Level 1:
$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn_{ti} + \pi_{3i} negmn_{ti} + \pi_{4i} posmn\_lag\_1_{ti} + \pi_{5i} negmn\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$

$\pi_{2i} = \beta_{20} + r_{2i}$
$\pi_{3i} = \beta_{30} + r_{3i}$
$\pi_{4i} = \beta_{40} + r_{4i}$
$\pi_{5i} = \beta_{50} + r_{5i}$

```
fit6 <- lmer(rse ~ rse_lag_1 + posmn + negmn + posmn_lag_1 + negmn_lag_1 +
    (posmn + negmn + rse_lag_1 + posmn_lag_1 + negmn_lag_1 | id), data = dat)

tab6 <- table_fun(fit6)

options(knitr.kable.NA = '')
tab6 %>% 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", 9, 10) %>%
  group_rows("Random", 5, 8) %>%
  add_header_above(c(" " = 1, "Model Q5" = 2))
```

	Model Q5	
	b	CI
Fixed		
Intercept	3.60	[2.92, 4.05]
rse_lag_1	0.35	[0.31, 0.44]
posmn	0.42	[0.29, 0.47]
negmn	-0.83	[-0.95, -0.73]
Random		
posmn_lag_1	-0.08	[-0.16, 0.02]
negmn_lag_1	0.05	[-0.08, 0.20]
τ_{00}	1.35	[0.50, 2.69]
τ_{11}	0.14	[0.10, 0.23]
Fit		
τ_{22}	0.09	[0.03, 0.23]
τ_{33}	0.04	[0.01, 0.08]
τ_{44}	0.11	[0.04, 0.22]
τ_{55}	0.05	[0.03, 0.14]
R^2_m	0.43	
R^2_c	0.67	

Part A

With current positive events controlled, is there any residual effect of the previous day's positive events on self-esteem? There is a positive effect of residual positive events on self-esteem (b = 0.42; 95% CI [0.29, 0.47]). The previous day's positive events predict higher self esteem.

Part B

With current negative events controlled, is there any residual effect of the previous day's negative events on self-esteem?

There is a negative effect of residual negative events on self-esteem (b = -0.83; 95% CI -0.83). The previous day's negative events predict lower self esteem.

Question 7

Lastly, take the model from the previous problem and add the standardized CES-D depression measure (you will need to create this) as a predictor in each Level 2 equation:

Level 1:
$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn_{ti} + \pi_{3i} negmn_{ti} + \pi_{4i} posmn\_lag\_1_{ti} + \pi_{5i} negmn\_lag\_1_{ti} + \epsilon_{ti}$
Level 2:
$\pi_{0i} = \beta_{00} + \beta_{01} cesd\_z_i + r_{0i}$
$\pi_{1i} = \beta_{10} + r_{1i}$
$\pi_{2i} = \beta_{20} + r_{2i}$
$\pi_{3i} = \beta_{30} + r_{3i}$
$\pi_{4i} = \beta_{40} + r_{4i}$
$\pi_{5i} = \beta_{50} + r_{5i}$

```
dat <- dat %>% group_by(id) %>%
  mutate(cesd_z = ifelse(is.na(cesd_z) == T, max(cesd_z, na.rm = T), cesd_z))
fit7 <- lmer(rse ~ rse_lag_1* + posmn*cesd_z + negmn*cesd_z + posmn_lag_1*cesd_z + negmn_lag_1*cesd_z +
    (posmn + negmn + rse_lag_1 + posmn_lag_1 + negmn_lag_1 | id), data = dat)
tab7 <- table_fun(fit7)
```

Part A

Increase the number of iterations used by the estimation algorithm.

```
cl1 <- lmerControl(optimizer = "bobyqa",
                   optCtrl=list(maxiter=1e9))

fit7a <- update(fit7, control = cl1)
tab7a <- table_fun(fit7a)
```

Part B

Try a different algorithm.

```
cl2 <- lmerControl(optimizer = "optimx",
                  optCtrl=list(method="nlminb",maxiter=1e9))

fit7b <- update(fit7, control = cl2)
tab7b <- table_fun(fit7b)
```

Part C

Simply the model by removing random effects (i.e., setting Level 2 variances to 0).

```
fit7c <- lmer(rse ~ rse_lag_1* + posmn*cesd_z + negmn*cesd_z + posmn_lag_1*cesd_z +
                negmn_lag_1*cesd_z + (1 | id), data = dat)
tab7c <- table_fun(fit7c)
```

```r
tab7 %>% mutate(model = "Full") %>%
  full_join(tab7a %>% mutate(model = "Increased Iter")) %>%
  full_join(tab7b %>% mutate(model = "Optimx")) %>%
  full_join(tab7c %>% mutate(model = "No RE")) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, model, est, sep = ".") %>%
  spread(key = tmp, value = value) %>%
  mutate(type = factor(type, levels = c("Fixed Parts", "Random Parts", "Model Terms"))) %>%
  arrange(type) %>%
  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", 1,14) %>%
  group_rows("Fit", 21, 22) %>%
  group_rows("Random", 15, 20) %>%
  add_header_above(c(" " = 1, "Full" = 2, "Increased Iter" = 2,
                     "Optimx" = 2, "No RE" = 2))
```

	Full		Increased Iter		Optimx		No RE	
	b	CI	b	CI	b	CI	b	CI
Fixed								
cesd_z	-0.37	[-0.58, 0.01]	-0.37	[-0.87, 0.06]	-0.40	[-0.87, -0.15]	-0.37	[-0.74, -0.10]
cesd_z:negmn	0.01	[-0.08, 0.16]	0.01	[-0.09, 0.16]	0.07	[-0.04, 0.24]	0.01	[-0.15, 0.16]
cesd_z:negmn_lag_1	0.07	[-0.12, 0.13]	0.07	[-0.04, 0.13]	0.07	[-0.05, 0.19]	0.07	[-0.04, 0.19]
cesd_z:posmn_lag_1	-0.15	[-0.26, -0.04]	-0.15	[-0.27, -0.11]	-0.15	[-0.20, -0.10]	-0.15	[-0.26, -0.05]
Intercept	3.13	[2.83, 3.79]	3.13	[2.62, 3.60]	3.04	[2.54, 3.45]	3.13	[2.77, 3.74]
negmn	-0.81	[-0.99, -0.72]	-0.81	[-0.89, -0.67]	-0.81	[-0.93, -0.66]	-0.81	[-0.86, -0.69]
negmn_lag_1	0.04	[-0.04, 0.14]	0.04	[-0.07, 0.24]	0.07	[-0.01, 0.15]	0.04	[-0.10, 0.18]
posmn	0.89	[0.33, 1.07]	0.89	[0.52, 1.16]	1.14	[0.71, 1.36]	0.89	[0.36, 1.09]
posmn:cesd_z	0.08	[-0.13, 0.34]	0.08	[-0.26, 0.44]	-0.04	[-0.17, 0.16]	0.08	[-0.04, 0.36]
posmn_lag_1	-0.07	[-0.15, -0.03]	-0.07	[-0.20, 0.08]	-0.05	[-0.12, 0.03]	-0.07	[-0.21, 0.05]
rse_lag_1	0.42	[0.30, 0.46]	0.42	[0.35, 0.49]	0.43	[0.37, 0.52]	0.42	[0.36, 0.48]
rse_lag_1:cesd_z	0.02	[-0.05, 0.08]	0.02	[-0.04, 0.12]	0.02	[-0.05, 0.11]	0.02	[-0.05, 0.11]
rse_lag_1:posmn	-0.08	[-0.11, 0.02]	-0.08	[-0.13, -0.02]	-0.13	[-0.17, -0.06]	-0.08	[-0.12, 0.01]
rse_lag_1:posmn:cesd_z	0.01	[-0.03, 0.06]	0.01	[-0.05, 0.06]	0.03	[0.00, 0.06]	0.01	[-0.03, 0.04]
Random								
τ_{00}	0.82	[0.48, 1.79]	0.82	[0.38, 2.25]	0.19	[0.17, 0.28]	0.82	[0.47, 2.26]
τ_{11}	0.11	[0.07, 0.23]	0.11	[0.06, 0.27]			0.11	[0.05, 0.23]
τ_{22}	0.12	[0.07, 0.34]	0.12	[0.04, 0.15]			0.12	[0.04, 0.57]
τ_{33}	0.03	[0.02, 0.07]	0.03	[0.02, 0.07]			0.03	[0.03, 0.06]
τ_{44}	0.10	[0.06, 0.19]	0.10	[0.02, 0.13]			0.10	[0.03, 0.13]
τ_{55}	0.09	[0.03, 0.23]	0.09	[0.03, 0.29]			0.09	[0.06, 0.21]
R^2_c	0.70		0.70		0.63		0.70	
R^2_m	0.50		0.50		0.48		0.50	

There is only one moderating effect of depression score: 1 SD change in depression is associated with a -0.15 * posmn decrease (95 % CI [-0.26, -0.04]) in self-esteem. In other words, positive events only benefit the self-esteem of those who are less depressed than average.

```
crossing(
  cesd_z = c(-1,0,1),
  posmn = 0,
  negmn = 0,
  posmn_lag_1 = seq(min(dat$posmn_lag_1, na.rm = T), max(dat$posmn_lag_1, na.rm = T),1),
  negmn_lag_1 = 0,
  rse_lag_1 = 0
) %>%
  mutate(pred = predict(fit7, newdata = ., re.form = NA),
         cesd = mapvalues(cesd_z, c(-1, 0, 1), c("-1 SD", "0 SD", "+1 SD"))) %>%
  ggplot(aes(x = posmn_lag_1, y = pred, color = cesd)) +
    geom_line() +
    theme_classic()
```