HW 1

Kavya Mudiam

4/9/2021

#PART 1 *I'm not sure if I'm supposed to only have timepoints 0-2 or 0-8*

```
d <- data %>%
  pivot_longer(cols = g3_fall:g5_spring,
               names_to = "term",
               values_to = "score") %>%
  select(-c(distid, scid)) %>%
  mutate(timepoint = term,
         timepoint=replace(timepoint, timepoint=="g3_fall", 0),
         timepoint=replace(timepoint, timepoint=="g3_winter", 1),
         timepoint=replace(timepoint, timepoint=="g3_spring", 2),
         timepoint=replace(timepoint, timepoint=="g4_fall", 3),
         timepoint=replace(timepoint, timepoint=="g4_winter", 4),
         timepoint=replace(timepoint, timepoint=="g4_spring", 5),
         timepoint=replace(timepoint, timepoint=="g5_fall", 6),
         timepoint=replace(timepoint, timepoint=="g5_winter", 7),
         timepoint=replace(timepoint, timepoint=="g5_spring", 8),
         timepoint = as.numeric(timepoint),
         grade = substr(term, start = 1, stop = 2),
         term = as.factor(term),
         #grade = sub("_....","  ", term),
         grade = as.factor(grade),
         sid = as.factor(sid))
```

conditional = predictor with something other than TIME (i.e., grade) #Question 2, Part A ##Unconditional growth model with random intercepts and parallel slopes

```
model1 <- lmer(score ~ 1 + timepoint + (1|sid), data = d)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ 1 + timepoint + (1 | sid)
##    Data: d
##
## REML criterion at convergence: 1390288
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1647 -0.6309  0.0014  0.6320  4.6756
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sid      (Intercept) 152.47   12.348
##  Residual              37.52    6.125
## Number of obs: 202500, groups:  sid, 22500
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 1.905e+02  8.606e-02  2213.8
## timepoint   3.673e+00  5.272e-03   696.8
##
## Correlation of Fixed Effects:
##           (Intr)
## timepoint -0.245
```

##Conditional growth model with random intercepts, parallel slopes, and grade-level fixed effects

```
model2 <- lmer(score ~ 1 + timepoint + grade + (1|sid), data = d)
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ 1 + timepoint + grade + (1 | sid)
##    Data: d
##
## REML criterion at convergence: 1363261
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6021 -0.6227 -0.0007  0.6216  4.7437
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sid      (Intercept) 153.05   12.371
##  Residual              32.28    5.682
## Number of obs: 202500, groups:  sid, 22500
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 188.79121    0.08672  2177.1
## timepoint     6.17704    0.01546   399.4
## gradeg4      -8.16802    0.05576  -146.5
## gradeg5     -16.69188    0.09780  -170.7
##
## Correlation of Fixed Effects:
##           (Intr) timpnt gradg4
## timepoint -0.178
## gradeg4    0.049 -0.832
## gradeg5    0.113 -0.949  0.877
```

##Unconditional growth model with random intercepts and random slopes

```
model3 <- lmer(score ~ 1 + timepoint + (1 + timepoint|sid), data= d) #Error
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00614329 (tol = 0.002, component 1)
```

```
summary(model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ 1 + timepoint + (1 + timepoint | sid)
##    Data: d
##
## REML criterion at convergence: 1359385
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8707 -0.6132 -0.0005  0.6148  4.2624
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  sid      (Intercept) 99.490   9.974
##           timepoint    1.242   1.114    0.38
##  Residual             28.201   5.310
## Number of obs: 202500, groups:  sid, 22500
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 1.905e+02  6.997e-02  2723.0
## timepoint   3.673e+00  8.723e-03   421.1
##
## Correlation of Fixed Effects:
##           (Intr)
## timepoint 0.174
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00614329 (tol = 0.002, component 1)
```

##Conditional growth model with random intercepts, random slopes, and grade-level fixed effects

```
model4 <- lmer(score ~ 1 + timepoint + grade + (1 + timepoint|sid), data = d)
summary(model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ 1 + timepoint + grade + (1 + timepoint | sid)
##    Data: d
##
## REML criterion at convergence: 1321862
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9013 -0.6063 -0.0018  0.6063  4.3282
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  sid      (Intercept) 101.752  10.087
##           timepoint     1.342   1.158   0.33
##  Residual              22.221   4.714
## Number of obs: 202500, groups:  sid, 22500
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 188.79121    0.07082  2665.6
## timepoint     6.17704    0.01497   412.5
## gradeg4      -8.16802    0.04626  -176.6
## gradeg5     -16.69188    0.08114  -205.7
##
## Correlation of Fixed Effects:
##           (Intr) timpnt gradg4
## timepoint  0.007
## gradeg4    0.050 -0.713
## gradeg5    0.115 -0.813  0.877
```

#PART B *Compare the performance of the four models you fit in the previous section. Which model displays the best fit to the data? Make a determination and provide a brief write up, using evidence to justify your selection.*
Given that the AIC and BIC are lowest for Model 4, it appears to be the best fit.

```
anova(model1, model2, model3, model4)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: d
## Models:
## model1: score ~ 1 + timepoint + (1 | sid)
## model2: score ~ 1 + timepoint + grade + (1 | sid)
## model3: score ~ 1 + timepoint + (1 + timepoint | sid)
## model4: score ~ 1 + timepoint + grade + (1 + timepoint | sid)
##        npar     AIC     BIC  logLik deviance   Chisq Df Pr(>Chisq)
## model1    4 1390284 1390325 -695138  1390276
## model2    6 1363253 1363314 -681620  1363241 27035.6  2  < 2.2e-16 ***
## model3    6 1359386 1359447 -679687  1359374  3866.7  0
## model4    8 1321858 1321939 -660921  1321842 37532.2  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#PART C - interpretation Provide a brief writeup interpreting the model you selected from above. Be sure to interpret both the fixed effects and random effects. I'm looking for a "plain English" description. It does not necessarily need to be APA style, but plain English and APA are also not mutually exclusive, so you could. Please make sure to also include confidence intervals in your interpretation.

```
confint(model4)
```

```
## Computing profile confidence intervals ...
```

```
##                   2.5 %      97.5 %
## .sig01        9.9867817  10.1886480
## .sig02        0.3162605   0.3457282
## .sig03        1.1446953   1.1720912
## .sigma        4.6975075   4.7304296
## (Intercept) 188.6523973 188.9300296
## timepoint     6.1476895   6.2063883
## gradeg4      -8.2586821  -8.0773529
## gradeg5     -16.8509185 -16.5328461
```

Computing profile confidence intervals …
2.5 % 97.5 % .sig01 9.9867817 10.1886480 .sig02 0.3162605 0.3457282 .sig03 1.1446953 1.1720912 .sigma 4.6975075 4.7304296 (Intercept) 188.6523973 188.9300296 timepoint 6.1476895 6.2063883 gradeg4 -8.2586821 -8.0773529 gradeg5 -16.8509185 -16.5328461

```
#Interpretation
*The intercept displays that on average students scored 188.79 points on the assessment
in the fall of Grade 3 ,95% CI: [188.65, 188.93]. The SD of random effect suggests that
the scores varied by 10.09 points per student, 95% CI: [9.99, 10.19]). Between each time
point, students improved 6.18 points on average, 95% CI: [6.15, 6.21], and the SD of 1.1
6 points, 95% CI: [1.14, 1.17], suggests that this was the variance between students ove
r time. During Grade 4, students lost 8.26 points on average, 95%  CI: -8.26, -8.08, and
16.69 points during Grade 5,95%  CI: [-16.85, -16.53].
```

Plot the predicted values for student ID's 1-1-1, 1-1-2, and 1-1-3, relative to their observed data points. Use facet wrapping to place them all in the same plot. The end result should look similar to the below, which shows this relation for student IDS 1-1-4, 1-1-5, and 1-1-6. Note that my plot has some styling added to it which you can feel

free to ignore (I just can’t help myself).