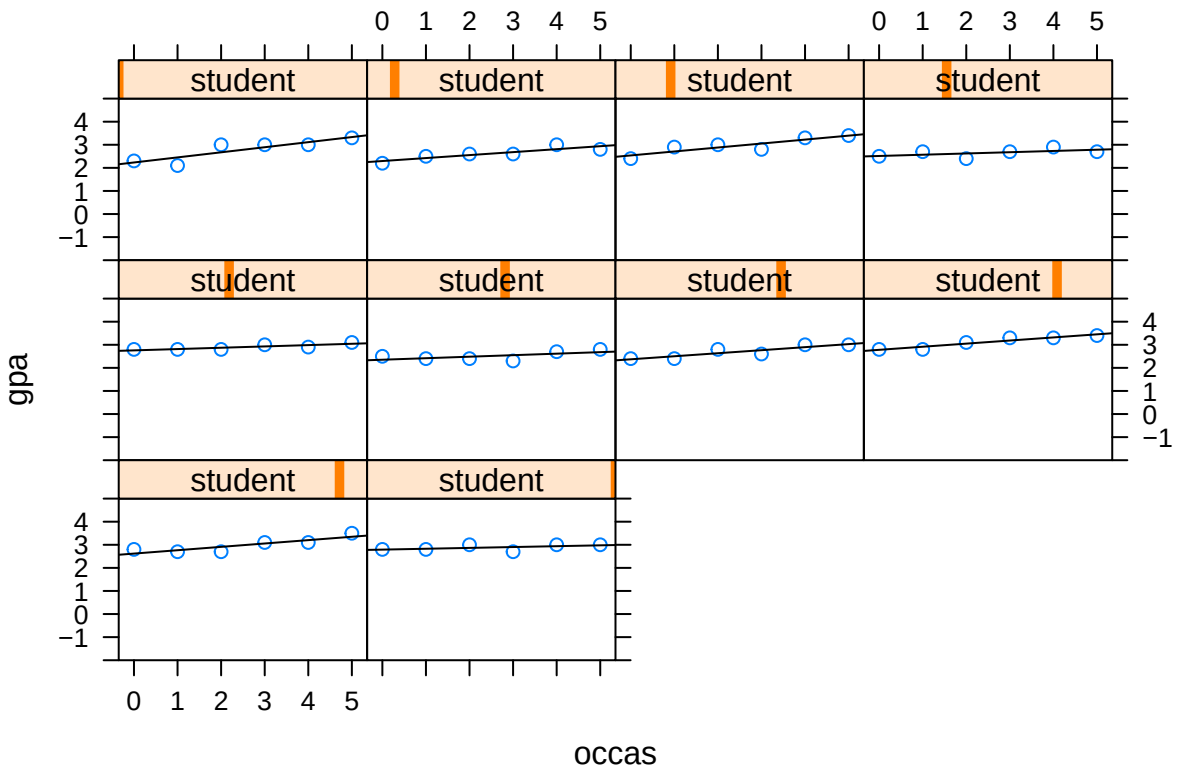
Assignment 2

Ayato Tanemura (atan524, 825159894)

Contents

Question 1

1



GPA and Occasion appears linear in six consecutive occasions.

ID 1 - 10: GPA tends to increase gradually from occasion 0 to occasion 5.

Generally there is considerable changes in GPA over years. And it is a linear trajectory for the identified change.

2

Unconditional means model

```
> Linear mixed-effects model fit by REML
>   Data: gpa
>       AIC      BIC    logLik
>   925.531 940.7988 -459.7655
>
> Random effects:
>  Formula: ~1 | student
>         (Intercept) Residual
> StdDev:     0.23903 0.312394
>
> Fixed effects:  gpa ~ 1
>             Value  Std.Error   DF  t-value p-value
```

```
> (Intercept) 2.865 0.01915729 1000 149.5514       0
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -3.65035984 -0.54961236  0.06029375  0.63563934  2.57357120
>
> Number of Observations: 1200
> Number of Groups: 200
```

- The composite model is :

GPA = 2.87 + e

- Estimate of fixed effect :

The initial status of GPA at occasion 0 is 2.87 (p-vale = 0) at 0.01 level of significance.

Unconditional growth model

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: gpa
>        AIC      BIC    logLik
>   258.2342 288.7747 -123.1171
>
> Random effects:
>  Formula: ~occas | student
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev     Corr
> (Intercept) 0.21179280 (Intr)
> occas       0.06685297 -0.096
> Residual    0.20588332
>
> Fixed effects:  gpa ~ occas
>                 Value   Std.Error  DF   t-value p-value
> (Intercept) 2.5992143 0.018326395 999 141.82900       0
> occas       0.1063143 0.005874944 999  18.09622       0
>  Correlation:
>       (Intr)
> occas -0.345
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -3.27468164 -0.53810705 -0.01278367  0.53269416  3.19453714
>
> Number of Observations: 1200
> Number of Groups: 200
```

- The composite model:

Level1: GPA = a + b * Occas + j

Level2: a = 2.60 + y_0i, b = 0.11 + y_1i

GPA = 2.60 + 0.11 * Occas + e (With composite residual : e = y_0i + y_1i * Occas + j)

- Estimate of fixed effect

1. The estimated initial GPA is 2.60 (p-vale = 0) at 0.01 level of significance.
2. The rate of change at the occasion 0 is 0.11 (p-vale = 0) at 0.01 level of significance.

3

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: gpa
>        AIC      BIC    logLik
>   245.9189 286.6395 -114.9595
>
> Random effects:
>  Formula: ~occas | student
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev     Corr
> (Intercept) 0.20840692 (Intr)
> occas       0.06527364 -0.14
> Residual    0.20588321
>
> Fixed effects:  gpa ~ sex * occas
>                 Value  Std.Error  DF  t-value p-value
> (Intercept) 2.5595489 0.02632912 998 97.21361  0.0000
> sex1        0.0755532 0.03633764 198  2.07920  0.0389
> occas       0.0911278 0.00840123 998 10.84696  0.0000
> sex1:occas  0.0289266 0.01159480 998  2.49479  0.0128
>  Correlation:
>            (Intr) sex1   occas
> sex1       -0.725
> occas      -0.380  0.276
> sex1:occas  0.276 -0.380 -0.725
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -3.26582864 -0.54107063 -0.01516501  0.54114011  3.22799961
>
> Number of Observations: 1200
> Number of Groups: 200
```

- Estimate of fixed effect:

1. The estimated initial GPA for the average male is 2.56 ($p < 0.01$) at 0.01 level of significance.
2. The estimated differential initial GPA between male and female is 0.076 ($p > 0.01$) at 0.01 level of significance.
3. The estimated rate of change in GPA for an average male is 0.29 ($p < 0.01$) at 0.01 level of significance.
4. The estimated differential in the rate of change in gender is 0.029. However, the estimate differential in the rate of change in GPA between male and female is indistinguishable ($p > 0.01$).

- Variance components

```
> student = pdLogChol(occas)
>             Variance    StdDev     Corr
> (Intercept) 0.043433442 0.20840692 (Intr)
> occas       0.004260648 0.06527364 -0.14
> Residual    0.042387897 0.20588321
```

Level 1 (within person variance) gets the estimate of 0.042.

Level 2 (between person variance) receives the estimate of 0.043 for the initial status and 0.0043 for the rate of change.

- Pseudo R2 statistics:

```
> student = pdLogChol(occas)
>             Variance   StdDev     Corr
> (Intercept) 0.04485619 0.21179280 (Intr)
> occas       0.00446932 0.06685297 -0.096
> Residual    0.04238794 0.20588332
```

```
> student = pdLogChol(occas)
>             Variance    StdDev     Corr
> (Intercept) 0.043433442 0.20840692 (Intr)
> occas       0.004260648 0.06527364 -0.14
> Residual    0.042387897 0.20588321
```

Pseudo R2 = (0.043433442 - 0.04485619) / 0.043433442 = -0.0328

-> 3.28% of the between-person variability in EMPL is associated with linear time.

4

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: gpa
>        AIC     BIC    logLik
>   238.7753 289.676 -109.3876
>
> Random effects:
>  Formula: ~occas | student
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev     Corr
> (Intercept) 0.20088516 (Intr)
> occas       0.06527345 -0.144
> Residual    0.20588323
>
> Fixed effects:  gpa ~ sex * occas + highgpa * occas
>                    Value  Std.Error  DF   t-value p-value
> (Intercept)    2.2758566 0.09423960 997 24.149685  0.0000
> sex1           0.0835949 0.03559781 197  2.348316  0.0198
> occas          0.0925278 0.03080142 997  3.004012  0.0027
> highgpa        0.0935466 0.02989492 197  3.129179  0.0020
> sex1:occas     0.0288869 0.01163484 997  2.482794  0.0132
> occas:highgpa -0.0004617 0.00977090 997 -0.047247  0.9623
>  Correlation:
```

```
>               (Intr) sex1   occas  highgp sx1:cc
> sex1          -0.267
> occas         -0.389  0.104
> highgpa       -0.962  0.072  0.374
> sex1:occas     0.104 -0.389 -0.267 -0.028
> occas:highgpa  0.374 -0.028 -0.962 -0.389  0.072
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -3.15955152 -0.54676739 -0.02154116  0.53719616  3.24751829
>
> Number of Observations: 1200
> Number of Groups: 200
```

- The effect on initial status and rate of change

1. The estimated initial GPA for the average male is 2.28 ($p < 0.01$).
2. The estimated differential in initial GPA between gender, controlling for GPA in high school is 0.084 ($p < 0.05$) at 0.05 level of significance.
3. The estimated rate of change in GPA for an average male, controlling for GPA in high school is 0.093 ($p < 0.01$) at 0.01 level of significance.
4. The estimated differential in initial GPA for difference in GPA controlling for gender at the initial stage is 0.094 ($p < 0.01$) at 0.01 level of significance.
5. The estimated differential in the rate of change in gender is 0.029. However, the estimated differential in the rate of change in GPA between gender is indistinguishable from 0 ($p > 0.01$).
6. The estimated differential in the rate of change of GDP in high school is -0.00046. However, the estimate differential in the rate of change in GPA of GPA in high school is indistinguishable ($p > 0.01$).

5

Let's say the unconditional growth model without predictor is model A, the unconditional growth model with the gender predictor is model B, and the unconditional growth model with the gender and GPA in high school predictors is model C.

The outcomes of AIC and BIC are the following;

AIC & BIC

model A: 258.2342, 288.7747

model B: 245.9189, 286.6395

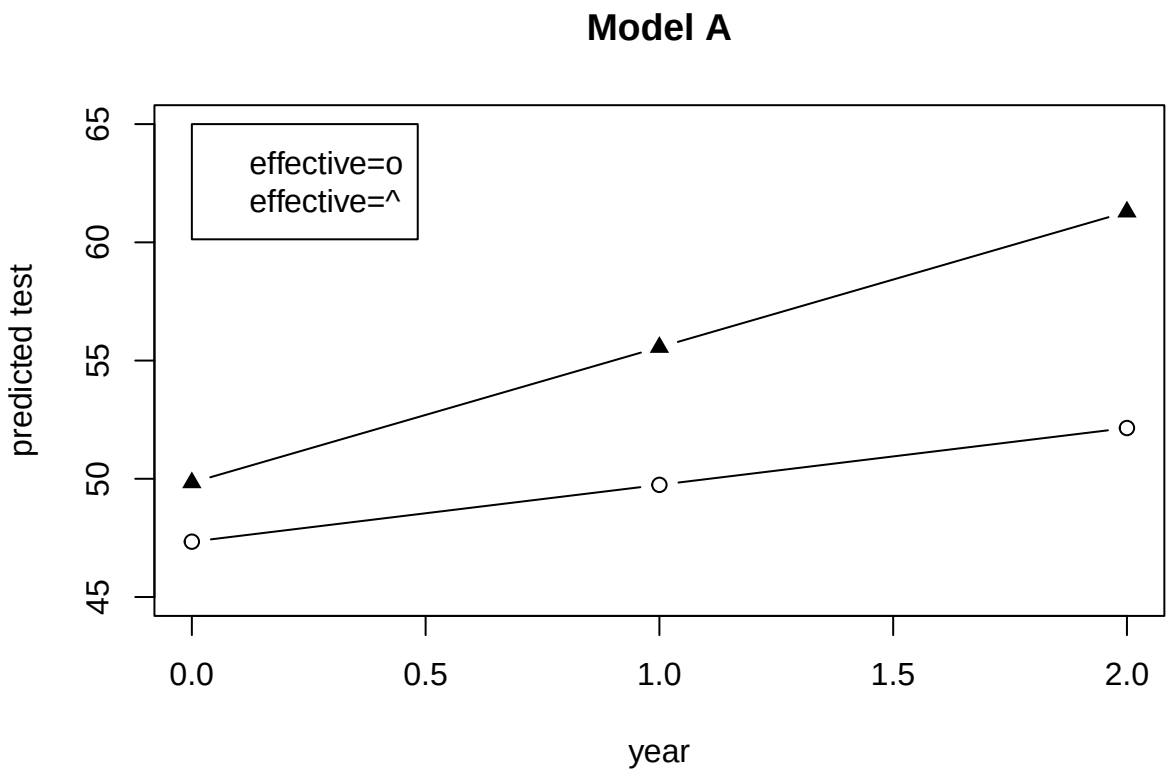
model C: 238.7753, 289.676

According to the outcome as above, AIC and BIC show the different result. Since AIC shows the bigger differences between two closet value than BIC, AIC would be used as the result.

Based on AIC, model C is the best model to explain the change in GPAs over time since AIC is the smallest.

Question 2

1



As you can see from the visualization, both categories are gradually increasing over the time.

To compare between the effectiveness 0 and 1, effectiveness 1 has higher initial math score and larger rate of change than effectiveness 0.

2

```
> Linear mixed-effects model fit by REML
>   Data: math
>       AIC      BIC    logLik
>   193897 193921.5 -96945.48
>
> Random effects:
>  Formula: ~1 | id
>         (Intercept) Residual
> StdDev:    5.523092   8.8375
>
> Fixed effects:  test ~ 1
>                Value  Std.Error    DF t-value p-value
> (Intercept) 52.94465 0.08075362 17340 655.632       0
>
> Standardized Within-Group Residuals:
```

```
>        Min         Q1        Med         Q3        Max
> -4.3186225 -0.5685095 -0.1096028  0.4357381  5.3466575
>
> Number of Observations: 26010
> Number of Groups: 8670
```

- Fixed effect:

The initial status of test score at the first occasion is 52.94 (p = 0) at 0.01 level of significance.

- Variance components

```
> id = pdLogChol(1)
>             Variance StdDev
> (Intercept) 30.50454 5.523092
> Residual    78.10140 8.837500
```

Level 1 (within person variance) gets the estimate of 78.10.

Level 2 (between person variance) receives the estimate of 30.50.

- ICC

ICC = 30.50 / (30.50 + 78.10) = 0.28

28% variation on test score is attributable to differences among students.

3

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: math
>        AIC      BIC    logLik
>   189288.7 189337.7 -94638.33
>
> Random effects:
>  Formula: ~time | id
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev   Corr
> (Intercept) 5.996266 (Intr)
> time        2.108674 -0.097
> Residual    7.466787
>
> Fixed effects:  test ~ time
>                Value  Std.Error    DF  t-value p-value
> (Intercept) 48.71361 0.09750184 17339 499.6173       0
> time         4.23105 0.06106083 17339  69.2923       0
>  Correlation:
>      (Intr)
> time -0.564
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
```

```
> -4.46708805 -0.33153438 -0.06521128  0.29050378  6.37602629
>
> Number of Observations: 26010
> Number of Groups: 8670
```

- Rate of change

The estimated rate of change in test score for an average student is 4.23 ($p < 0.01$) at 0.01 level of significance.

So we can interpret this estimate: there is a increase in test score over the three occasions with a 4.23 score of increase each year.

4

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: math
>        AIC      BIC    logLik
>   187170.1 187235.4 -93577.04
>
> Random effects:
>  Formula: ~time | id
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev   Corr
> (Intercept) 5.866091 (Intr)
> time        1.309675 -0.427
> Residual    7.466785
>
> Fixed effects:  test ~ effective * time
>                   Value  Std.Error    DF  t-value p-value
> (Intercept)    47.33965 0.14399383 17338 328.7617       0
> effective       2.49784 0.19415106  8668  12.8655       0
> time            2.40374 0.08710244 17338  27.5967       0
> effective:time  3.32202 0.11744274 17338  28.2863       0
>  Correlation:
>                (Intr) effctv time
> effective      -0.742
> time           -0.637  0.472
> effective:time  0.472 -0.637 -0.742
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -4.60457887 -0.35146914 -0.06488285  0.28756448  6.42412623
>
> Number of Observations: 26010
> Number of Groups: 8670
```

- Fixed effect

1. The estimated initial test score for the average student of effective 0 is 47.33 ($p < 0.01$) at 0.01 level of significance.
2. The estimated differential in initial test score between effective 0 and 1 is 2.50 ($p < 0.01$) at 0.01 level of significance.

3. the estimated rate of change in test score for an average student of effective 0 is 2.40 ($p < 0.01$) at 0.01 level of significance.

4. The estimated differential in the rate of change in test score between effective 0 and 1 is 3.32 ($p < 0.01$) at 0.01 level of significance.

- variance components

```
> id = pdLogChol(time)
>             Variance  StdDev   Corr
> (Intercept) 34.411029 5.866091 (Intr)
> time         1.715248 1.309675 -0.427
> Residual    55.752882 7.466785
```

Level 1 (within person variance) gets the estimate of 55.75.

Level 2 (between person variance) receives the estimate of 34.41 for the initial status and 1.72 for the rate of change.

5

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: math
>        AIC      BIC    logLik
>   187169.4 187251.1 -93574.72
>
> Random effects:
>  Formula: ~time | id
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev   Corr
> (Intercept) 5.863402 (Intr)
> time        1.304308 -0.426
> Residual    7.466776
>
> Fixed effects:  test ~ effective * time + ses * time
>                   Value  Std.Error    DF  t-value p-value
> (Intercept)    47.33120 0.14404419 17337 328.5880  0.0000
> effective       2.49789 0.19412051  8667  12.8677  0.0000
> time            2.40938 0.08712920 17337  27.6530  0.0000
> ses             0.22758 0.12346845  8667   1.8433  0.0653
> effective:time  3.32199 0.11741928 17337  28.2917  0.0000
> time:ses       -0.15187 0.07468338 17337  -2.0335  0.0420
>  Correlation:
>                (Intr) effctv time   ses    effct:
> effective      -0.741
> time           -0.637  0.472
> ses            -0.032  0.000  0.020
> effective:time  0.472 -0.637 -0.741  0.000
> time:ses        0.020  0.000 -0.032 -0.637  0.000
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -4.58376764 -0.35157943 -0.06581068  0.28939500  6.43088933
```

```
>
> Number of Observations: 26010
> Number of Groups: 8670
```

- Fixed Effect

1. The estimated initial test score for the average teacher's effectiveness 0, controlling for ses is 47.33 ($p < 0.01$) at 0.01 level of significance.
2. The estimated differential in initial test score between effectiveness 0 and 1, controlling for ses is 2.50 ($p < 0.01$) at 0.01 level of significance.
3. The estimated rate of change in test score for an average teacher's effectiveness 0, controlling for ses is 2.41 ($p < 0.01$) at 0.01 level of significance.
4. The estimated differential in initial test score for 1 score difference in ses controlling for effectiveness at the initial stage is 0.23 ($p < 0.1$) at 0.1 level of significance.
5. The estimated differential in the rate of change in test score between effectiveness 0 and 1 is 3.31 ($p < 0.01$) at 0.01 level of significance.
6. The estimated differential in the rate of change in test score of ses is 0.15 ($p < 0.05$) at 0.05 level of significance.

We conclude that students who have got teacher's effectiveness 1 get higher test score initially than teacher's effective 0 and the rate of change for teacher's effective 1 is higher than effective 0.

We also conclude that ses is positively associated with the early test score but negatively associated with the rate of change in the test score

In the first year, student who have got a higher score for ses tend to get a higher test score. But they have a slower rate of increase in the test score over time.

6

```
> Linear mixed-effects model fit by maximum likelihood
>   Data: math
>        AIC      BIC    logLik
>   187171.6 187245.1 -93576.79
>
> Random effects:
>  Formula: ~time | id
>  Structure: General positive-definite, Log-Cholesky parametrization
>             StdDev   Corr
> (Intercept) 5.864723 (Intr)
> time        1.309680 -0.426
> Residual    7.466783
>
> Fixed effects:  test ~ ses + effective * time
>                   Value  Std.Error    DF  t-value p-value
> (Intercept)    47.33714 0.14402570 17338 328.6715  0.0000
> ses             0.06772 0.09520131  8667   0.7113  0.4769
> effective       2.49786 0.19413550  8667  12.8666  0.0000
> time            2.40374 0.08710411 17338  27.5962  0.0000
> effective:time  3.32202 0.11744500 17338  28.2858  0.0000
```

```
>  Correlation:
>                (Intr) ses    effctv time
> ses            -0.025
> effective      -0.741  0.000
> time           -0.637  0.000  0.472
> effective:time  0.472  0.000 -0.637 -0.742
>
> Standardized Within-Group Residuals:
>         Min          Q1         Med          Q3         Max
> -4.60112083 -0.35132406 -0.06472555  0.28710666  6.42511668
>
> Number of Observations: 26010
> Number of Groups: 8670
```

- Fixed Effect

1. The estimated initial test score for teacher's effectiveness 0, controlling for ses is 47.33 ($p < 0.01$) at 0.01 level of significance.
2. The estimated differential in initial test score for 1 score difference in ses controlling for effectiveness at the initial stage is 0.0677 ($p > 0.1$) at non-significant level.
3. The estimated differential in initial test score between effectiveness 0 and 1 controlling for ses is 2.498 ($p < 0.01$) at 0.01 level of significance.
4. The estimated rate of change in test score for 1 score difference in ses is 2.404 ($p < 0.01$) at 0.01 level of significance.
5. The estimated differential in the rate of change in test score of ses is 3.32 ($p < 0.01$) at 0.01 level of significance.

- Variance Components

```
> id = pdLogChol(time)
>             Variance  StdDev   Corr
> (Intercept) 34.394978 5.864723 (Intr)
> time         1.715261 1.309680 -0.426
> Residual    55.752851 7.466783
```

Level 1 (within person variance) gets the estimate of 55.75.

Level 2 (between person variance) receives the estimate of 34.39 for the initial status and 1.72 for the rate of change.

7

Let's say the unconditional growth model without any predictors is model C, the unconditional growth model with the effectiveness predictor is model D, the unconditional growth model with the effectiveness and ses predictors is model E, and the unconditional growth model with the ses predictor only for initial status and the effectiveness predictor for time is model F.

The outcomes of AIC and BIC are the following;

AIC & BIC

model C: 189288.7, 189337.7

model D: 187170.1, 187235.4

model E: 187169.4, 187251.1

model F: 187171.6, 187245.1

To compare the numbers by AIC, model C is the largest and model E is the smallest, however, in the BIC case, the model with the largest number is the same but the model with the smallest number is model D.

In my perspective, I would rather use AIC to select the best model since the gap between the smalles number and the second smallest number is larger than BIC.

Therefore, model E is the most appropriate to use at the end.