Assignment 2 - Language Development in ASD - Part 1 - Explaining development

Studygroup 4

11/9/2020

#CLEAN markdown with #small comments on choises and results are made along the way while the full report can be seen in the attached document.

See the document here

Loading usefull packages

```
pacman::p_load(tidyverse, purrr, MuMIn, lmerTest)
```

Loading the clean dataframe from last portfolio

```
df<- read.csv("autism_df.csv")
#removing x
df[,1] <- NULL
```

Exploring Participants

```
#exploring the format of the variables
str(df)
```

```
## 'data.frame':    372 obs. of  20 variables:
##  $ SUBJ         : int  1 1 1 1 1 1 2 2 2 2 ...
##  $ ADOS1        : int  0 0 0 0 0 0 13 13 13 13 ...
##  $ nonverbalIQ1 : int  28 28 28 28 28 28 34 34 34 34 ...
##  $ verbalIQ1    : int  14 14 14 14 14 14 27 27 27 27 ...
##  $ Socialization1: int  108 108 108 108 108 108 85 85 85 85 ...
##  $ VISIT        : int  1 2 3 4 5 6 1 2 3 4 ...
##  $ Ethnicity    : chr  "White" "White" "White" "White" ...
##  $ Diagnosis    : chr  "TD" "TD" "TD" "TD" ...
##  $ Gender       : chr  "M" "M" "M" "M" ...
##  $ Age          : num  19.8 23.9 27.7 32.9 35.9 ...
##  $ ADOS         : int  0 NA NA NA 0 NA 13 NA NA NA ...
##  $ nonverbalIQ  : int  28 NA NA 33 NA 42 34 NA NA 49 ...
##  $ verbalIQ     : int  14 NA NA NA NA 44 27 NA NA NA ...
##  $ Socialization: int  108 110 109 102 107 100 85 105 77 75 ...
##  $ MOT_MLU      : num  3.62 3.86 4.32 4.42 5.21 ...
##  $ CHI_MLU      : num  1.25 1.01 1.56 2.25 3.24 ...
##  $ types_MOT    : int  378 403 455 533 601 595 317 307 351 335 ...
##  $ types_CHI    : int  14 18 97 133 182 210 146 171 262 200 ...
##  $ tokens_MOT   : int  1835 2160 2149 2260 2553 2586 1428 1270 1445 1286 ...
##  $ tokens_CHI   : int  139 148 255 321 472 686 461 562 983 674 ...
```

```
#Mutating variables
df <- df %>%
  mutate(SUBJ = as.factor(SUBJ),
         #VISIT = as.factor(VISIT),
         Gender = as.factor(Gender),
         Diagnosis = as.factor(Diagnosis),
         Ethnicity = as.factor(Ethnicity)
        )
#Checking the difference in each condition by "splitting"/grouping Td from ASD
df %>%
  split(df$Diagnosis) %>%
  map(summary)
```

```
## $ASD
##      SUBJ          ADOS1          nonverbalIQ1      verbalIQ1       Socialization1
## 2      :  6   Min.   : 0.00   Min.   :13.00   Min.   : 8.00   Min.   : 64.00
## 4      :  6   1st Qu.:11.00   1st Qu.:25.00   1st Qu.:11.00   1st Qu.: 69.00
## 5      :  6   Median :14.00   Median :27.00   Median :16.00   Median : 76.00
## 6      :  6   Mean   :14.11   Mean   :26.89   Mean   :17.58   Mean   : 77.20
## 7      :  6   3rd Qu.:17.00   3rd Qu.:30.00   3rd Qu.:24.50   3rd Qu.: 85.25
## 18     :  6   Max.   :21.00   Max.   :42.00   Max.   :33.00   Max.   :105.00
## (Other):140
##      VISIT                 Ethnicity    Diagnosis Gender       Age
## Min.   :1.000   White          :130   ASD:176   F: 26   Min.   :18.77
## 1st Qu.:2.000   African American: 12   TD :  0   M:150   1st Qu.:36.88
## Median :3.000   White/Latino    : 12                     Median :42.87
## Mean   :3.443   Lebanese        :  6                      Mean   :43.17
## 3rd Qu.:5.000   White/Asian     :  6                      3rd Qu.:50.03
## Max.   :6.000   Bangladeshi     :  5                      Max.   :62.40
##                 (Other)         :  5                      NA's   :4
##      ADOS         nonverbalIQ      verbalIQ       Socialization
## Min.   : 0.0   Min.   :13.00   Min.   : 8.00   Min.   : 38.00
## 1st Qu.:11.0   1st Qu.:27.00   1st Qu.:11.75   1st Qu.: 68.00
## Median :14.0   Median :31.00   Median :19.50   Median : 74.00
## Mean   :13.8   Mean   :33.15   Mean   :22.62   Mean   : 77.34
## 3rd Qu.:17.0   3rd Qu.:40.50   3rd Qu.:30.00   3rd Qu.: 85.00
## Max.   :25.0   Max.   :50.00   Max.   :50.00   Max.   :116.00
## NA's   :117   NA's   :89     NA's   :116    NA's   :1
##     MOT_MLU        CHI_MLU         types_MOT       types_CHI
## Min.   :1.856   Min.   :0.000   Min.   : 74.0   Min.   :  0.00
## 1st Qu.:3.222   1st Qu.:1.012   1st Qu.:284.2   1st Qu.:  9.00
## Median :3.699   Median :1.370   Median :340.0   Median : 50.50
## Mean   :3.657   Mean   :1.642   Mean   :338.7   Mean   : 77.21
## 3rd Qu.:4.100   3rd Qu.:2.158   3rd Qu.:397.8   3rd Qu.:144.75
## Max.   :5.380   Max.   :4.302   Max.   :585.0   Max.   :307.00
## NA's   :10     NA's   :10     NA's   :10     NA's   :10
##    tokens_MOT     tokens_CHI
## Min.   : 209   Min.   :   0.0
## 1st Qu.:1388   1st Qu.:  58.0
## Median :1805   Median : 196.0
## Mean   :1780   Mean   : 293.6
## 3rd Qu.:2220   3rd Qu.: 482.0
## Max.   :3182   Max.   :1293.0
```

```
##  NA's   :10     NA's   :10
##
## $TD
##      SUBJ         ADOS1          nonverbalIQ1      verbalIQ1       Socialization1
## 1      :  6   Min.   : 0.000   Min.   :17.00   Min.   :13.00   Min.   : 84.0
## 3      :  6   1st Qu.: 0.000   1st Qu.:24.00   1st Qu.:17.00   1st Qu.: 96.0
## 8      :  6   Median : 0.000   Median :27.00   Median :19.00   Median :102.0
## 9      :  6   Mean   : 0.949   Mean   :25.93   Mean   :20.14   Mean   :100.5
## 10     :  6   3rd Qu.: 1.000   3rd Qu.:29.00   3rd Qu.:22.00   3rd Qu.:104.0
## 12     :  6   Max.   :15.000   Max.   :32.00   Max.   :33.00   Max.   :115.0
## (Other):160                    NA's   :1       NA's   :1
##      VISIT                  Ethnicity    Diagnosis Gender       Age
## Min.   :1.000    White           :190   ASD:  0   F: 36    Min.   :18.07
## 1st Qu.:2.000    Asian           :  6   TD :196   M:160    1st Qu.:23.96
## Median :3.000    African American:  0                      Median :30.88
## Mean   :3.434    Bangladeshi     :  0                      Mean   :30.60
## 3rd Qu.:5.000    Bangledeshi     :  0                      3rd Qu.:36.33
## Max.   :6.000    Hispanic        :  0                      Max.   :45.07
##                  (Other)         :  0                      NA's   :6
##      ADOS          nonverbalIQ       verbalIQ      Socialization
## Min.   : 0.000   Min.   :17.0   Min.   :13.00   Min.   : 59.0
## 1st Qu.: 0.000   1st Qu.:28.0   1st Qu.:18.00   1st Qu.: 97.0
## Median : 0.000   Median :39.0   Median :28.00   Median :102.0
## Mean   : 1.273   Mean   :35.8   Mean   :29.29   Mean   :101.8
## 3rd Qu.: 1.000   3rd Qu.:44.0   3rd Qu.:40.00   3rd Qu.:107.0
## Max.   :15.000   Max.   :50.0   Max.   :50.00   Max.   :125.0
## NA's   :130      NA's   :101    NA's   :133     NA's   :2
##     MOT_MLU         CHI_MLU          types_MOT       types_CHI
## Min.   :2.776   Min.   :0.5584   Min.   :178.0   Min.   :  7.00
## 1st Qu.:3.805   1st Qu.:1.5577   1st Qu.:305.5   1st Qu.: 72.25
## Median :4.117   Median :2.3202   Median :363.5   Median :131.00
## Mean   :4.150   Mean   :2.3064   Mean   :368.6   Mean   :129.31
## 3rd Qu.:4.472   3rd Qu.:2.9095   3rd Qu.:418.8   3rd Qu.:181.00
## Max.   :5.744   Max.   :4.3648   Max.   :601.0   Max.   :298.00
## NA's   :10      NA's   :10       NA's   :10      NA's   :10
##    tokens_MOT      tokens_CHI
## Min.   : 584    Min.   :  16.0
## 1st Qu.:1503    1st Qu.: 254.2
## Median :1844    Median : 435.0
## Mean   :1879    Mean   : 475.7
## 3rd Qu.:2264    3rd Qu.: 669.8
## Max.   :3077    Max.   :1294.0
## NA's   :10      NA's   :10
```

REPORT IN DOCS

Hypothesis 1: The child's MLU changes: i) over time, ii) according to diagnosis

Ploting the data

```
#Plotting Mean length utterance for all visits for each diagnosis in linear
df %>%
  ggplot(aes(x = VISIT, y = CHI_MLU, color = Diagnosis)) +
    geom_point(stat='summary') +
    geom_smooth(method = "lm")
```

Warning: Removed 20 rows containing non-finite values (stat_summary).

No summary function supplied, defaulting to `mean_se()`

`geom_smooth()` using formula 'y ~ x'

Warning: Removed 20 rows containing non-finite values (stat_smooth).
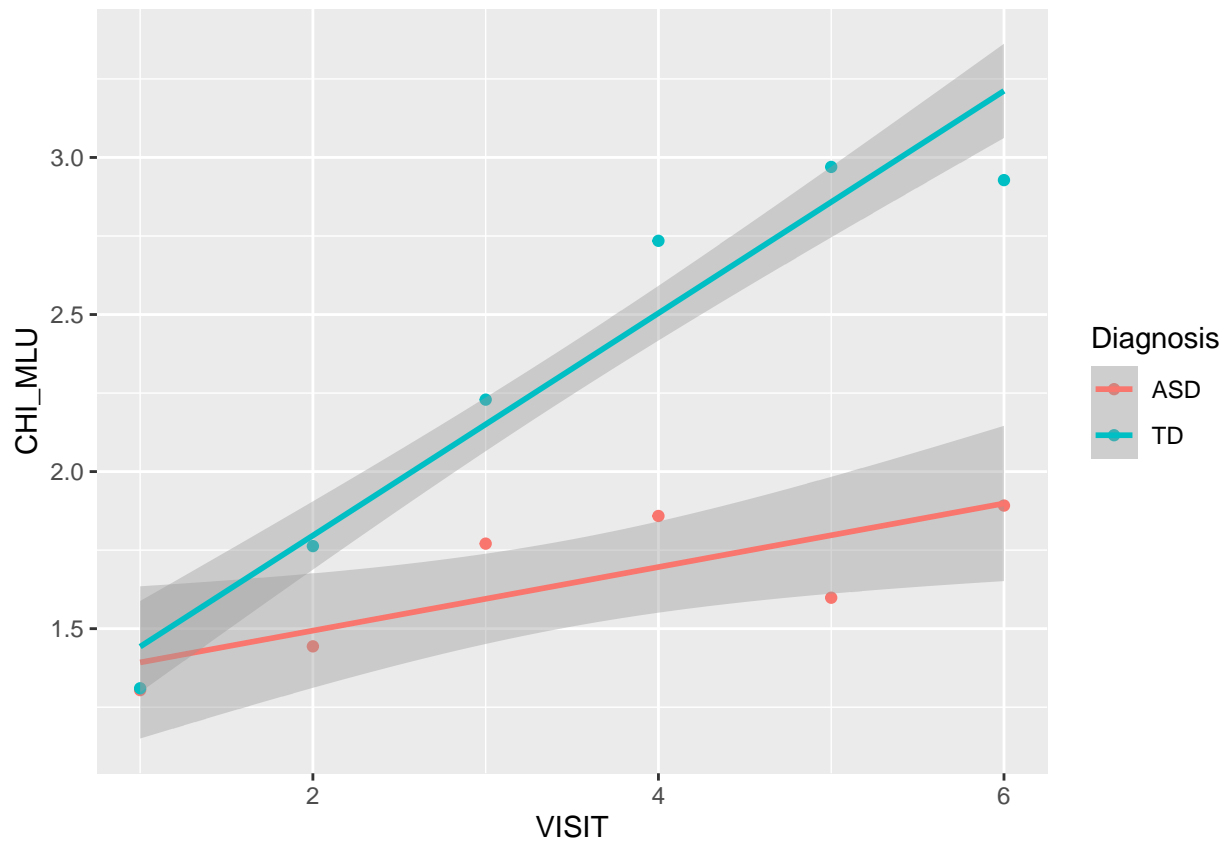
Making a model Fixed effect: Random effect: To account for independence, *linear mixed-effects model* is used

```
#null model
null_model <- lmerTest::lmer(CHI_MLU ~ 1+(1|SUBJ),
                     df,
                   REML = FALSE)

summary(null_model)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ 1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    831.0    842.6   -412.5    825.0      349
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.36375 -0.60029 -0.02655  0.59406  2.59982
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.4272   0.6536
##  Residual             0.4401   0.6634
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  1.99141    0.09089 60.92299   21.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m0 <- lmerTest::lmer(CHI_MLU ~ Gender + Age +(1+VISIT|SUBJ)+
                     VISIT*Diagnosis,
                   df,
                 REML = FALSE)
summary(m0)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ Gender + Age + (1 + VISIT | SUBJ) + VISIT * Diagnosis
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    562.5    601.0   -271.3    542.5      336
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.48854 -0.52868 -0.06189  0.44192  2.72904
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.30720  0.5543
##           VISIT       0.01103  0.1050   -0.21
##  Residual             0.15769  0.3971
## Number of obs: 346, groups:  SUBJ, 61
##
## Fixed effects:
##                   Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        1.61242    0.54091 87.69470   2.981  0.00372 **
## GenderM            0.21774    0.20444 58.78288   1.065  0.29120
## Age               -0.01742    0.01706 92.13547  -1.021  0.30972
```

```
## VISIT                0.17542    0.07590 115.07048   2.311  0.02261 *
## DiagnosisTD          -0.40862    0.27587  85.45712  -1.481  0.14223
## VISIT:DiagnosisTD     0.24850    0.03718  61.40133   6.685 7.84e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GendrM Age    VISIT  DgnsTD
## GenderM     -0.334
## Age         -0.917  0.010
## VISIT        0.818 -0.009 -0.935
## DiagnosisTD -0.832  0.045  0.781 -0.653
## VISIT:DgnTD  0.059 -0.001  0.023 -0.280 -0.278
```

```
#Age is highly correlated with visit. Therefore, it should be removed.

m1 <- lmerTest::lmer(CHI_MLU ~ VISIT*Diagnosis+Gender+
                     (1+VISIT|SUBJ),
                   df,
                 REML = FALSE)
summary(m1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ VISIT * Diagnosis + Gender + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    573.3    608.1   -277.7    555.3      343
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -2.48637 -0.53772 -0.08284  0.44456  2.72812
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.28752  0.5362
##           VISIT       0.01123  0.1060   -0.16
##  Residual             0.16062  0.4008
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##                   Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        1.11628    0.21332 65.49037   5.233 1.89e-06 ***
## VISIT              0.10050    0.02680 61.42377   3.750 0.000394 ***
## DiagnosisTD       -0.20611    0.16851 59.31840  -1.223 0.226140
## GenderM            0.21838    0.20321 59.94455   1.075 0.286825
## VISIT:DiagnosisTD  0.25330    0.03713 61.98037   6.822 4.36e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) VISIT  DgnsTD GendrM
## VISIT       -0.255
```

```
## DiagnosisTD -0.461   0.322
## GenderM     -0.821   0.001   0.059
## VISIT:DgnTD  0.183 -0.722 -0.447   0.000
```

```
# Gender has a high p -value. Let us try without it.

m2 <- lmerTest::lmer(CHI_MLU ~ VISIT*Diagnosis+
          (1+VISIT|SUBJ),
                  df,
                REML = FALSE)
summary(m2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    572.5    603.4   -278.2    556.5      344
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.48470 -0.53248 -0.08812  0.44179  2.73712
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.29420  0.5424
##           VISIT       0.01122  0.1059   -0.16
##  Residual             0.16063  0.4008
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##                   Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        1.30459    0.12273 59.16414  10.630 2.40e-15 ***
## VISIT              0.10046    0.02680 61.43186   3.749 0.000395 ***
## DiagnosisTD       -0.21693    0.16952 59.26133  -1.280 0.205657
## VISIT:DiagnosisTD  0.25331    0.03712 61.98846   6.823 4.35e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) VISIT  DgnsTD
## VISIT       -0.443
## DiagnosisTD -0.724  0.321
## VISIT:DgnTD  0.320 -0.722 -0.445
```

```
# Only with the interaction effect

m3 <- lmerTest::lmer(CHI_MLU ~ VISIT+(1+VISIT|SUBJ),
                  df,
                REML = FALSE)

summary(m3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ VISIT + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    605.4    628.6   -296.7    593.4      346
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.57451 -0.08056  0.42981  2.71573
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.30038  0.5481
##           VISIT       0.02683  0.1638   -0.24
##  Residual             0.16143  0.4018
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  1.19062    0.08535 59.06421   13.95  < 2e-16 ***
## VISIT        0.23346    0.02452 60.80205    9.52 1.17e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## VISIT -0.431
```

 # VISIT as the only predictor

Assessing the model

```
anova(m1, m2, m3)
```

```
## Data: df
## Models:
## m3: CHI_MLU ~ VISIT + (1 + VISIT | SUBJ)
## m2: CHI_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
## m1: CHI_MLU ~ VISIT * Diagnosis + Gender + (1 + VISIT | SUBJ)
##    npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m3    6 605.44 628.62 -296.72   593.44
## m2    8 572.46 603.37 -278.23   556.46 36.9792  2  9.334e-09 ***
## m1    9 573.32 608.09 -277.66   555.32  1.1441  1     0.2848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

 # m2 has the best AIC and BIC score. Furthermore, the anova tells us that m2 is significantly different

 # How about the R squared value?
r.squaredGLMM(m1)

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
##           R2m       R2c
## [1,] 0.3660161 0.8154333
```

```r
r.squaredGLMM(m2)
```

```
##           R2m       R2c
## [1,] 0.3571597 0.8150068
```

```r
r.squaredGLMM(m3)
```

```
##           R2m       R2c
## [1,] 0.1824133 0.8146647
```

```r
# We have chosen m2 as our model and now we want to see if this model is better than the null model
anova(m2, null_model)
```

```
## Data: df
## Models:
## null_model: CHI_MLU ~ 1 + (1 | SUBJ)
## m2: CHI_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## null_model    3 831.02 842.61 -412.51   825.02
## m2            8 572.46 603.37 -278.23   556.46 268.56  5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our model is significantly better than the null model.

Checking assumptions

Doesn't seem too good yet

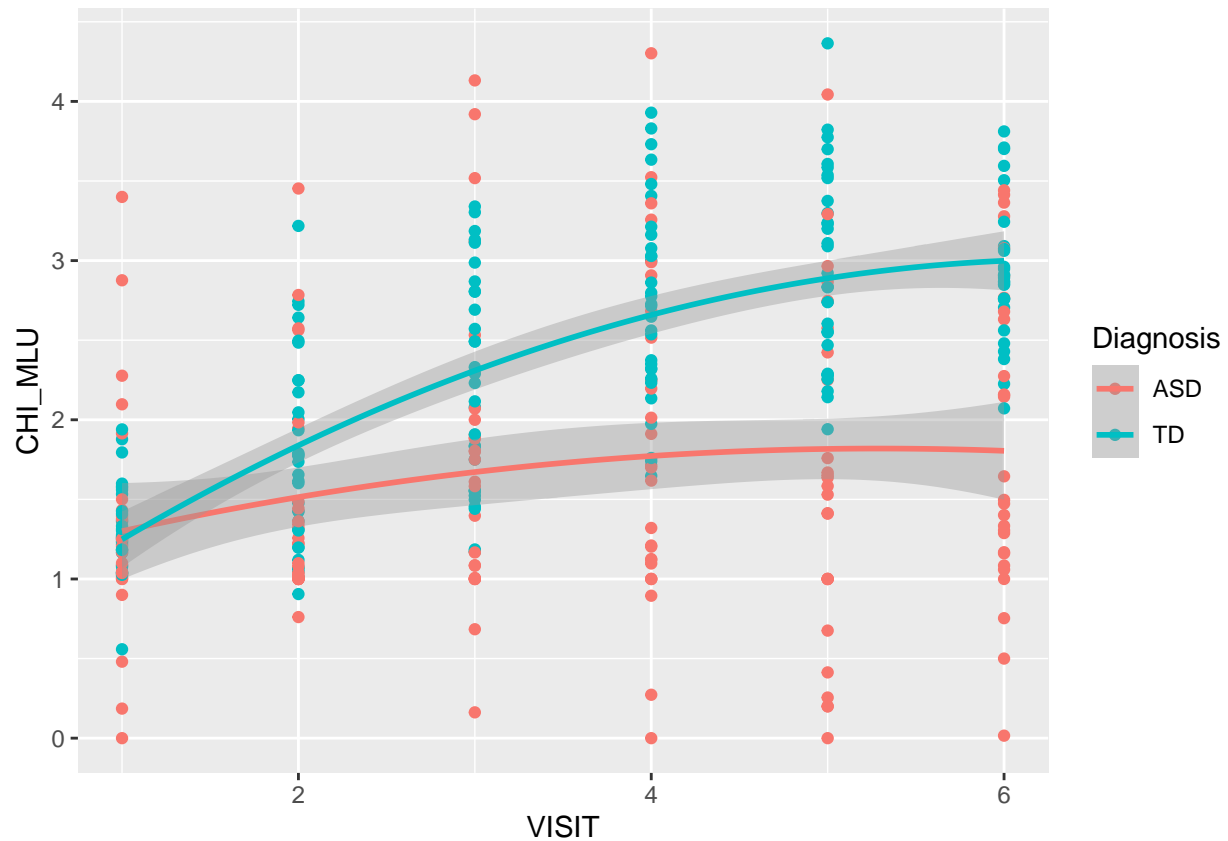**Creating growth models**

Visualizing quadratic

```r
qplot(VISIT,CHI_MLU, data = df, color = Diagnosis)+stat_smooth(method="lm", formula="y~poly(x,2)")
```

```
## Warning: Removed 20 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 20 rows containing missing values (geom_point).
```

Quadratic model

```
growth <- lmerTest::lmer(CHI_MLU ~ VISIT*Diagnosis+I(VISIT^2)+(1+I(VISIT^2)|SUBJ),
                         df,
                         REML = FALSE)
```

Assesing the quadratic model

Comparing models

```
anova(m2, growth)
```

```
## Data: df
## Models:
## m2: CHI_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
## growth: CHI_MLU ~ VISIT * Diagnosis + I(VISIT^2) + (1 + I(VISIT^2) |
## growth:      SUBJ)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m2         8 572.46 603.37 -278.23   556.46
## growth     9 549.46 584.23 -265.73   531.46 25.002  1  5.727e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

growth seems to be the best fit.

Plotting the actual CHI_MLU data against the predictions of the model fitted(model).

# Hypothesis 2: The parents's MLU changes: i) over time, ii) according to diagnosis

Ploting the data

```
#Plotting Mean length utterance for all visits for each diagnosis in linear
df %>%
  ggplot(aes(x = VISIT, y = MOT_MLU, color = Diagnosis)) +
    geom_point() +
    geom_smooth(method = "lm")
```

`geom_smooth()` using formula 'y ~ x'

Warning: Removed 20 rows containing non-finite values (stat_smooth).

Warning: Removed 20 rows containing missing values (geom_point).
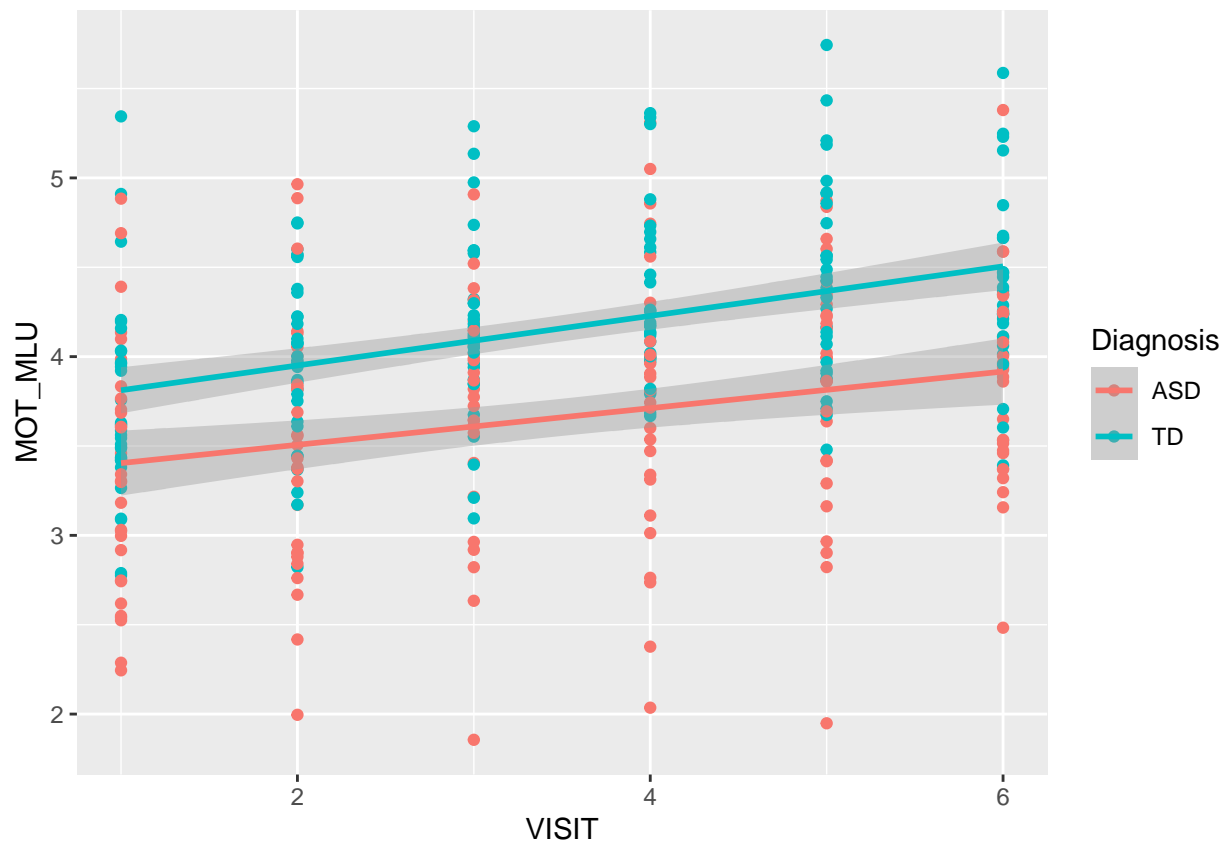
Making a model

```
#null model
mot_null <- lmerTest::lmer(MOT_MLU ~ 1+(1|SUBJ),
                     df,
                  REML = FALSE)

summary(mot_null)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: MOT_MLU ~ 1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    606.8    618.3   -300.4    600.8      349
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.08641 -0.55417 -0.00895  0.59362  2.82192
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.2217   0.4708
##  Residual             0.2335   0.4832
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  3.91165    0.06559 60.82484   59.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
mot_m1 <- lmerTest::lmer(MOT_MLU ~ VISIT*Diagnosis+
            (1+VISIT|SUBJ),
                   df,
                REML = FALSE)
summary(mot_m1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: MOT_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    513.5    544.4   -248.7    497.5      344
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.88746 -0.58770 -0.03689  0.53919  2.96482
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.32460  0.5697
##           VISIT       0.01097  0.1048   -0.69
##  Residual             0.14715  0.3836
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##                 Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)      3.31421    0.12529 59.76870  26.452  < 2e-16 ***
## VISIT            0.09922    0.02610 59.09554   3.801 0.000343 ***
## DiagnosisTD      0.35647    0.17304 59.84355   2.060 0.043752 *
```

12

```
## VISIT:DiagnosisTD  0.04022    0.03616 59.61514   1.112 0.270395
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) VISIT  DgnsTD
## VISIT       -0.755
## DiagnosisTD -0.724  0.547
## VISIT:DgnTD  0.545 -0.722 -0.755
```

```
# Only with the interaction effect

mot_m2 <- lmerTest::lmer(MOT_MLU ~ VISIT+Diagnosis+
            (1+VISIT|SUBJ),
                  df,
                REML = FALSE)

summary(mot_m2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: MOT_MLU ~ VISIT + Diagnosis + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    512.7    539.8   -249.4    498.7      345
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -2.89517 -0.59461 -0.02917  0.53790  3.00471
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.32943  0.5740
##           VISIT       0.01135  0.1065   -0.70
##  Residual             0.14723  0.3837
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  3.23810    0.10541 80.04973  30.720  < 2e-16 ***
## VISIT        0.12024    0.01823 59.52730   6.595 1.26e-08 ***
## DiagnosisTD  0.50193    0.11334 60.91617   4.429 4.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) VISIT
## VISIT       -0.627
## DiagnosisTD -0.566  0.003
```

```
mot_m3 <- lmerTest::lmer(MOT_MLU ~ VISIT+(1+VISIT|SUBJ),
                  df,
```

```
                      REML = FALSE)
```

```
summary(mot_m3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: MOT_MLU ~ VISIT + (1 + VISIT | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    527.4    550.6   -257.7    515.4      346
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.91830 -0.59032 -0.03879  0.56572  2.96334
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SUBJ     (Intercept) 0.35562  0.5963
##           VISIT       0.01125  0.1060   -0.59
##  Residual             0.14741  0.3839
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  3.50251    0.08934 60.22266   39.20  < 2e-16 ***
## VISIT        0.11991    0.01820 59.63373    6.59 1.28e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## VISIT -0.687
```

```
# VISIT as the only predictor
```

Assessing the model

```
anova(mot_null, mot_m1, mot_m2, mot_m3)
```

```
## Data: df
## Models:
## mot_null: MOT_MLU ~ 1 + (1 | SUBJ)
## mot_m3: MOT_MLU ~ VISIT + (1 + VISIT | SUBJ)
## mot_m2: MOT_MLU ~ VISIT + Diagnosis + (1 + VISIT | SUBJ)
## mot_m1: MOT_MLU ~ VISIT * Diagnosis + (1 + VISIT | SUBJ)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mot_null    3 606.76 618.35 -300.38   600.76
## mot_m3      6 527.44 550.62 -257.72   515.44 85.321  3  < 2.2e-16 ***
## mot_m2      7 512.71 539.75 -249.35   498.71 16.728  1  4.314e-05 ***
## mot_m1      8 513.48 544.39 -248.74   497.48  1.227  1      0.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Let us check r squared
r.squaredGLMM(mot_null)
```

```
##      R2m       R2c
## [1,]   0 0.4870039
```

```
r.squaredGLMM(mot_m1)
```

```
##          R2m       R2c
## [1,] 0.2292245 0.6753041
```

```
r.squaredGLMM(mot_m2)
```

```
##          R2m       R2c
## [1,] 0.2296138 0.6762255
```

```
r.squaredGLMM(mot_m3)
```

```
##          R2m       R2c
## [1,] 0.0924636 0.6748082
```

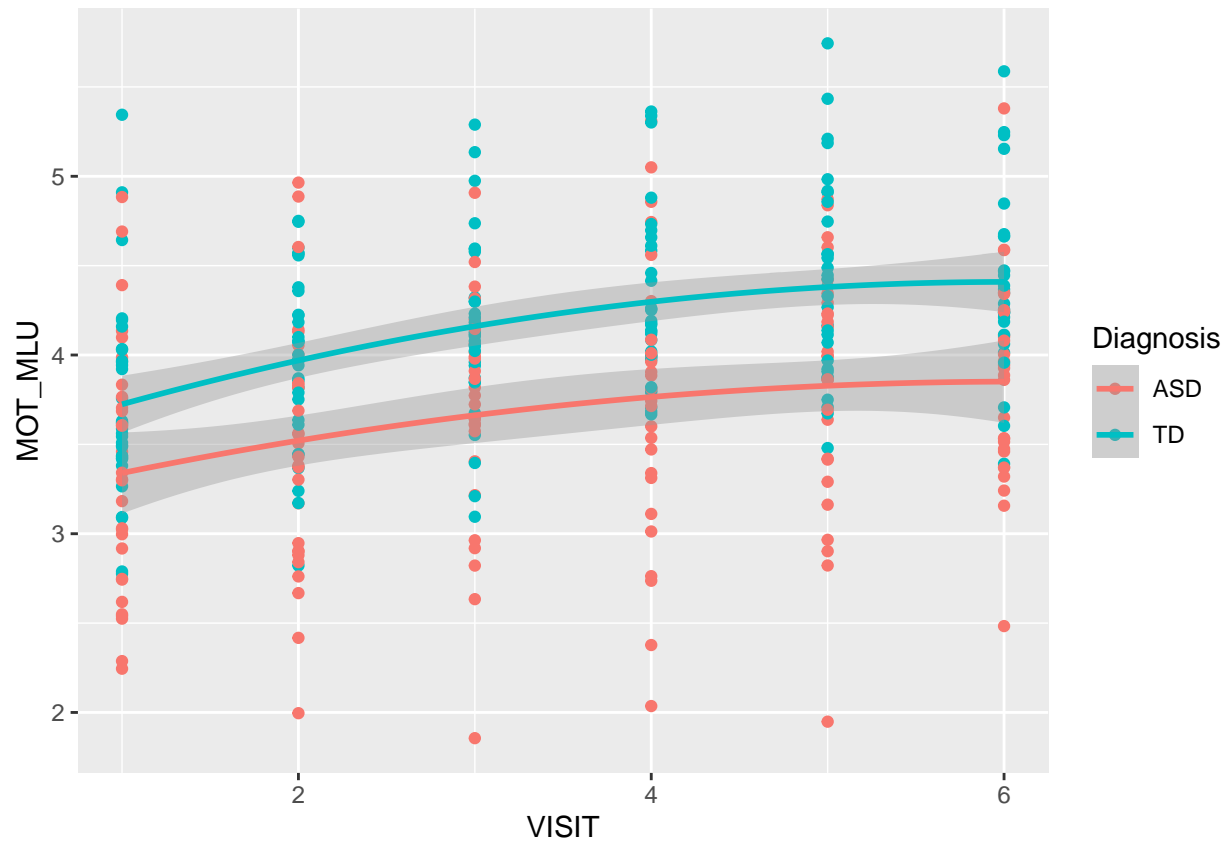m2 is the best model when comparing AIC, BIC and logLik scores.

Checking the model

Creating growth models

Visualizing quadratic

```
qplot(VISIT,MOT_MLU, data = df, color = Diagnosis)+stat_smooth(method="lm", formula="y~poly(x,2)")
```

```
## Warning: Removed 20 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 20 rows containing missing values (geom_point).
```

Quadratic Model

```
mot_growth <- lmerTest::lmer(MOT_MLU ~ Diagnosis+I(VISIT^2)+(1|SUBJ),
                        df,
                        REML = FALSE)

summary(mot_growth)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: MOT_MLU ~ Diagnosis + I(VISIT^2) + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    539.6    558.9   -264.8    529.6      347
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8005 -0.6047 -0.0262  0.5979  3.5663
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.1641   0.4051
##  Residual             0.1940   0.4404
## Number of obs: 352, groups:  SUBJ, 61
##
```

```
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 3.427e+00  8.761e-02 7.669e+01  39.117  < 2e-16 ***
## DiagnosisTD 4.974e-01  1.141e-01 6.084e+01   4.361 5.09e-05 ***
## I(VISIT^2)  1.512e-02  1.943e-03 2.919e+02   7.779 1.27e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) DgnsTD
## DiagnosisTD -0.686
## I(VISIT^2)  -0.331  0.005
```

Checking the quadratic model

Comparing models

```
anova(mot_m2, mot_growth)
```

```
## Data: df
## Models:
## mot_growth: MOT_MLU ~ Diagnosis + I(VISIT^2) + (1 | SUBJ)
## mot_m2: MOT_MLU ~ VISIT + Diagnosis + (1 + VISIT | SUBJ)
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mot_growth    5 539.56 558.88 -264.78   529.56
## mot_m2        7 512.71 539.75 -249.35   498.71 30.853  2  1.997e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

m2 seems to be the best fit. When plotting this model agianst the actual data, it can be seen that their is a linear trend.

- Excercise 4: Adding new variables *Your task now is to figure out how to best describe the children linguistic trajectory. The dataset contains a bunch of additional demographic, cognitive and clinical variables (e.g.verbal and non-verbal IQ). Try them out and identify the statistical models that best describes your data (that is, the children's MLU). Describe how you selected the best model and send the code to run the model to Victor and Byurakn.*

Making models

```
#Modelling the 4 additional predictors individually to Child MLU to check which one makes sense to incl
bm1 <- lmer(CHI_MLU ~ Socialization1 + (1|SUBJ), df, REML = F)
summary(bm1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ Socialization1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    808.8    824.2   -400.4    800.8      348
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.39210 -0.56995  0.01333  0.61927  2.55090
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.2610   0.5109
##  Residual             0.4404   0.6636
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##                Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   -0.542793   0.469812 60.436965  -1.155    0.252
## Socialization1 0.028335   0.005186 60.432777   5.463 9.31e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## Socializtn1 -0.987
```

```r
r.squaredGLMM(bm1)
```

```
##            R2m       R2c
## [1,] 0.1918285 0.4925739
```

```r
  #Socialization predicts okay
```

```r
bm2 <- lmer(CHI_MLU ~ ADOS1 + (1|SUBJ), df, REML = F)
summary(bm2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ ADOS1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    805.1    820.5   -398.5    797.1      348
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.38868 -0.55857  0.03541  0.61564  2.54803
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.2402   0.4901
##  Residual             0.4406   0.6638
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  2.418334   0.101510 59.798494  23.823  < 2e-16 ***
## ADOS1       -0.059331   0.009939 60.026622  -5.969 1.38e-07 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## ADOS1 -0.704
```

```r
r.squaredGLMM(bm2)
```

```
##           R2m       R2c
## [1,] 0.2151538 0.4920158
```

```r
  #Not good

bm3 <- lmer(CHI_MLU ~ nonverbalIQ1 + (1|SUBJ), df, REML = F)
summary(bm3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ nonverbalIQ1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    820.0    835.4   -406.0    812.0      348
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -2.42068 -0.60838 -0.04024  0.60845  2.63156
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.3306   0.5750
##  Residual             0.4400   0.6634
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   0.18595    0.48088 60.51504   0.387 0.700348
## nonverbalIQ1  0.06835    0.01794 60.59914   3.810 0.000327 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## nonverblIQ1 -0.985
```

```r
r.squaredGLMM(bm3)
```

```
##           R2m       R2c
## [1,] 0.1133578 0.4937446
```

```
  #Pretty good

bm4 <- lmer(CHI_MLU ~ verbalIQ1 + (1|SUBJ), df, REML = F)
summary(bm4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: CHI_MLU ~ verbalIQ1 + (1 | SUBJ)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    778.4    793.9   -385.2    770.4      348
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.62734 -0.63502 -0.05694  0.58760  2.97791
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SUBJ     (Intercept) 0.1299   0.3605
##  Residual             0.4398   0.6632
## Number of obs: 352, groups:  SUBJ, 61
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  0.379930   0.181448 62.004559   2.094   0.0404 *
## verbalIQ1    0.085574   0.009123 62.087824   9.380 1.65e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## verbalIQ1 -0.947
```

```
r.squaredGLMM(bm4)
```

```
##            R2m       R2c
## [1,] 0.3415312 0.4916978
```

```
  #Pretty good
```

Plotting simple models

Making max-models

Checking assumptions

Congrats! You made it to the end :D