BCP MLU Preliminary Analysis

Trevor KM Day

June 25, 2020
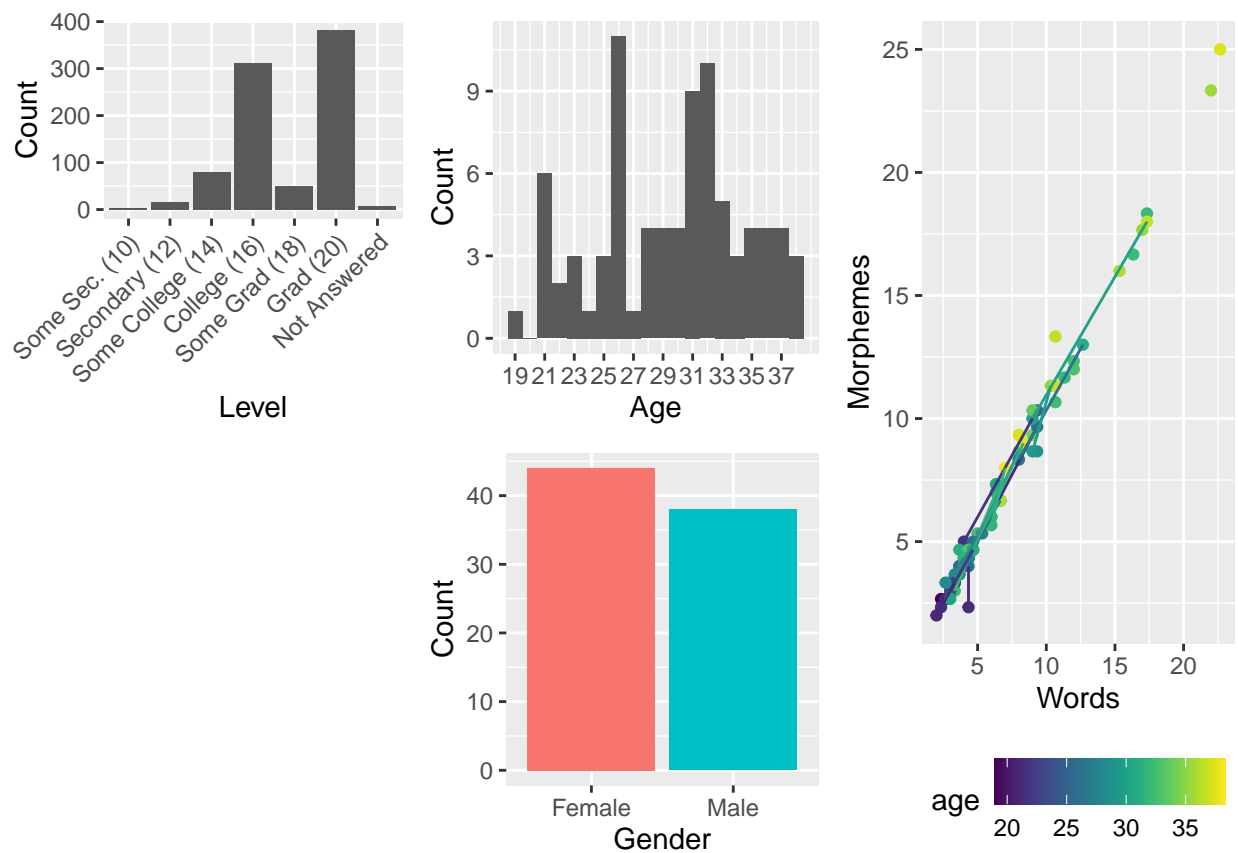
Setup

Packages: `tidyverse`, `lme4`, `MuMIn`, `viridis`, `gridExtra`.

Load data

Wordbank had only age, sex, and mother's education as demographics. Here, we show those for the available subjects, as well as the correlation between MLU3m and MLU3w.

```
## [1] 0.9941398
```

```
##        Age Mom Ed   LEX  SYN Part I PartII MLU3m MLU3w
## Mean 29.7   17.9 374.3 71.9  416.2   30.0   7.2   6.8
## SD    4.9    2.4 163.7 54.4  192.5   23.1   4.8   4.5
```

Analysis

Since MLU3m and MLU3w are so closely correlated, I use MLU3m for here on out.

So we see that control overall improves the models, and that SYN/Part II are better than LEX/Part I, but SYN doesn't meaningfully improve over Part II.

Reanalysis

However, it turns out SYN is a superset of Part II, so we can do a direct ANOVA for more information.

```r
# Demographic predictors
demo.predictors <- c("age", "sex", "educ_momed", "income_inr")

# All part I/lexical predictors
lexical.predictors <- c("action_words", "animals", "body_parts", "clothing",
                        "descriptive_words", "food_drink", "furniture_rooms",
                        "games_routines", "helping_verbs", "household",
                        "outside", "people", "places", "toys", "vehicles",
                        "sounds")

# Predictors in Part II, but also syntax
swap.predictors <- c("pronouns ", "quantifiers", "question_words", "time_words",
                     "word_endings_nouns", "connecting_words", "locations")

# Part II/syntax
partII.predictors <- c("word_endings_verbs", "word_forms_nouns",
                       "word_forms_verbs", "complexity")

formula.syntax_demo <- paste("MLU3m ~ ",
                             paste(c(demo.predictors, "(1|data_id)"),
                                   collapse = " + "))

formula.syntax_partII <- paste("MLU3m ~ ",
                               paste(c(demo.predictors, partII.predictors),
                                     collapse = " + "),
                               "+ (1|data_id)")

formula.syntax_SYN <- paste("MLU3m ~",
                            paste(c(demo.predictors, swap.predictors,
                                    partII.predictors),
                                  collapse = " + "),
                            "+ (1|data_id)")

syntax00 <- lmer(MLU3m ~ 1 + (1|data_id), data = wMLU, REML = TRUE)
syntax0  <- lmer(formula = formula.syntax_demo, data = wMLU, REML = TRUE)
syntax1  <- lmer(formula = formula.syntax_partII, data = wMLU, REML = TRUE)
```

```
syntax2  <- lmer(formula = formula.syntax_SYN, data = wMLU, REML = TRUE)

anova(syntax00, syntax0, syntax1, syntax2)
```

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


## Data: wMLU
## Models:
## syntax00: MLU3m ~ 1 + (1 | data_id)
## syntax0: MLU3m ~ age + sex + educ_momed + income_inr + (1 | data_id)
## syntax1: MLU3m ~ age + sex + educ_momed + income_inr + word_endings_verbs +
## syntax1:     word_forms_nouns + word_forms_verbs + complexity + (1 | data_id)
## syntax2: MLU3m ~ age + sex + educ_momed + income_inr + pronouns + quantifiers +
## syntax2:     question_words + time_words + word_endings_nouns + connecting_words +
## syntax2:     locations + word_endings_verbs + word_forms_nouns + word_forms_verbs +
## syntax2:     complexity + (1 | data_id)
##          npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## syntax00    3 490.32 497.54 -242.16   484.32
## syntax0    10 439.74 463.81 -209.87   419.74 64.5811  7  1.825e-11 ***
## syntax1    14 418.87 452.56 -195.43   390.87 28.8743  4  8.291e-06 ***
## syntax2    21 426.51 477.05 -192.25   384.51  6.3583  7     0.4986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Improvements:
# demo   >  intercept-only
# partII >  demo
# SYN    !> partII
```

Conclusion

Preliminary analysis with 82 subjects shows that MLU3m is differentially associated with syntax metrics over lexical metrics. However, the proposed metric is not significantly better than traditional MCDI Part II.

Analysis plan:

- Identify other variables and test them further against BCP82 sample.
- Once phenoscreening is done, use any variables identified above and rerun.

Package info

```
version
```

```
##                         _
## platform      x86_64-w64-mingw32
## arch          x86_64
## os            mingw32
## system        x86_64, mingw32
## status
## major         4
## minor         0.3
## year          2020
## month         10
## day           10
## svn rev       79318
## language      R
## version.string R version 4.0.3 (2020-10-10)
## nickname      Bunny-Wunnies Freak Out
```

```
installed.packages()[names(sessionInfo()$otherPkgs), "Version"]
```

```
##    gridExtra      viridis viridisLite        MuMIn         lme4       Matrix
##        "2.3"      "0.5.1"      "0.3.0"   "1.43.17"     "1.1-26"      "1.3-2"
##      forcats      stringr        dplyr        purrr        readr        tidyr
##      "0.5.0"      "1.4.0"      "1.0.3"      "0.3.4"      "1.4.0"      "1.1.2"
##       tibble      ggplot2    tidyverse
##      "3.0.5"      "3.3.3"      "1.3.0"
```