Section 6: Linear Growth Models

Josh Gilbert (adapted from past S43 TFs)

2023-09-26

Table of contents

1 Warmup

Let's check out the visualizer for our random slope models and work through the prompts as a group.

2 Goals for today

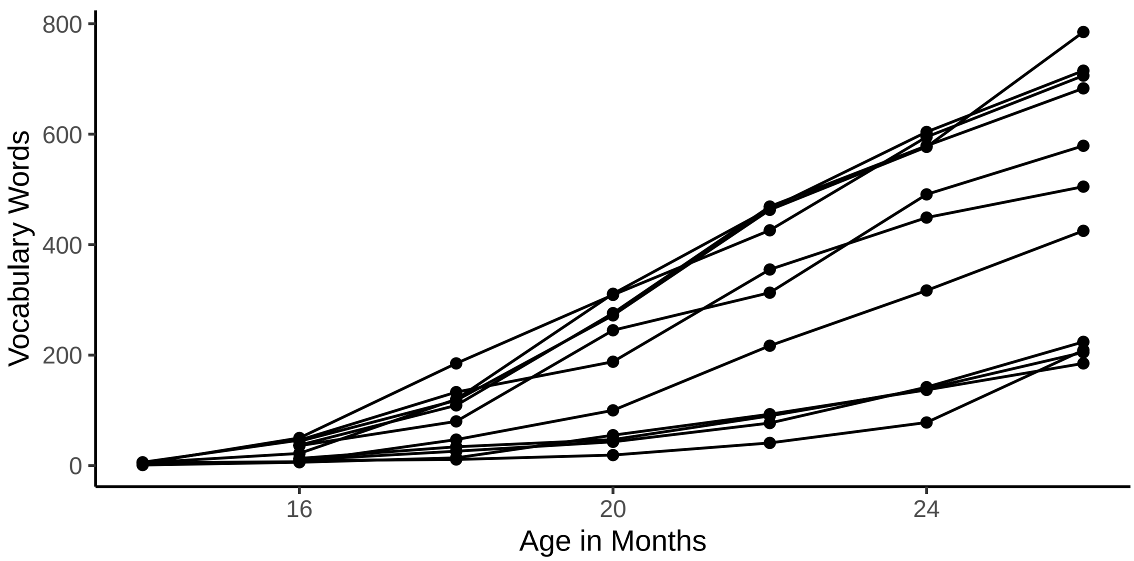
1. Review notation unconditional linear and quadratic growth models
2. Review quadratic growth model output in R
3. Review useful code for creating pretty plots

3 EDA: Empirical Growth Trajectories

Let's start with some EDA. Here, we're plotting the raw measurements from our data with "connect the dots" lines, not plotting fitted growth curve models. This is always the place to start!

```
plot_1 <- ggplot(data = dat, aes(x = age, y = vocab, group = pers)) +
  geom_point() +
  geom_line() +
  labs(x = "Age in Months",
       y = "Vocabulary Words")

plot_1
```



4 Activity 1: Growth Model Notation

4.1 (Unconditional) Linear growth model

$$vocab_{ti} = \pi_{0i} + \pi_{1i}a_{ti} + \epsilon_{ti}$$

$$\epsilon_{ti} \sim N(0, \sigma^2)$$

$$\pi_{0i} = \gamma_{00} + u_{0i}$$
$$\pi_{1i} = \gamma_{10} + u_{1i}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

Define the following in the above model:

- Level 1: time, measured in months
- Level 2: individual students
- a_{ti}: age in months at measurement point t for person i
- π_{0i}: an individual student's "initial" vocabulary level, i.e., # vocab words at $a_{ti} = 0$
- π_{1i}: an individual student's vocabulary growth rate per unit of time, i.e., expected # vocab words growth per month
- β_{00}: the "grand intercept" - mean initial vocab level for all students (when age is zero)
- β_{10}: the "grand slope" - mean growth rate per month for all students
- τ_{00}: extent of individual variation around the mean initial vocab level (variance of π_{0i})
- τ_{11}: extent of individual variation around the mean growth rate/curve (variance of π_{1i})
- $\rho = \frac{\tau_{10}}{\sqrt{\tau_{00}\tau_{11}}}$: correlation between initial status (intercepts u_{0i}) and rate of change (slopes u_{1i})
- ϵ_{ti}: residual for measurement of vocab at time t for individual i - how far a particular measurement is different from the model's prediction
- σ^2: variance of all of the residuals (how much of the data is left unexplained after our model?/how wrong is our model, on average (squared)?)

4.2 Fitting and Graphing the Linear Growth Model

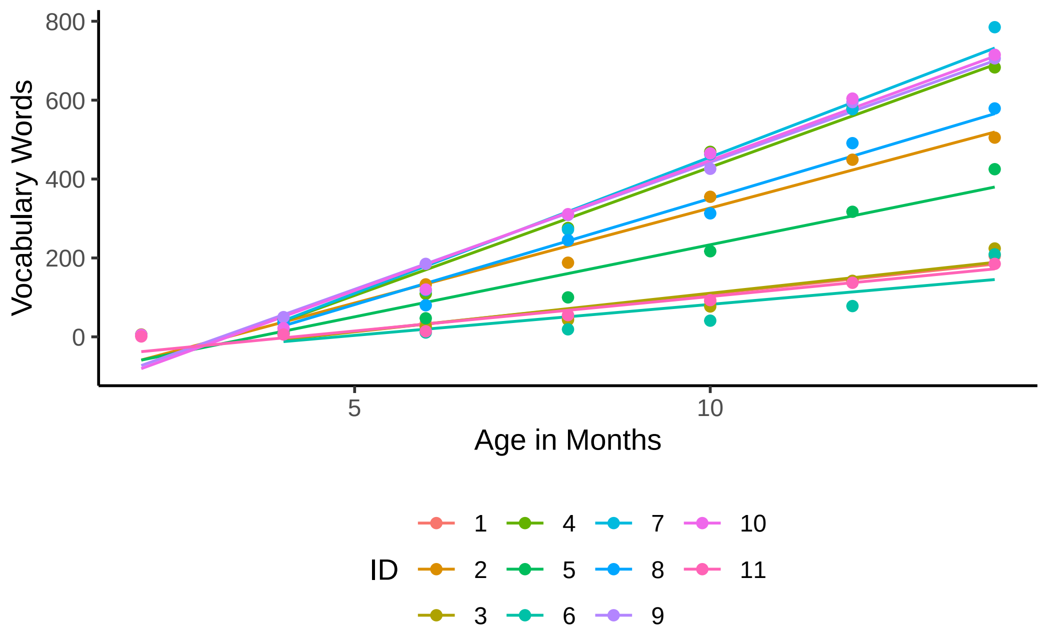
The code below fits and graphs the linear growth model. Let's take a moment to connect the output to the Greek and to the graphs, as a group.

```r
# fit the model
m0 <- lmer(vocab ~ 1 + age + (1 + age|pers),
           data = dat)
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00522975 (tol = 0.002, component 1)

# display the results
m0 |> display()
## lmer(formula = vocab ~ 1 + age + (1 + age | pers), data = dat)
##             coef.est coef.se
## (Intercept) -668.65   102.87
## age           43.16     6.79
##
## Error terms:
##  Groups   Name        Std.Dev. Corr
##  pers     (Intercept) 331.37
##           age          22.17   -1.00
##  Residual              36.73
## ---
## number of obs: 71, groups: pers, 11
## AIC = 765.6, DIC = 777
## deviance = 765.3

# get the fitted values
dat <- dat |>
  mutate(linear_fit = fitted(m0))

# plot and compare to the empirical data
ggplot(dat, aes(x = age12, y = linear_fit, color = factor(pers))) +
  geom_line() +
  geom_point(aes(y = vocab)) +
  labs(x = "Age in Months",
       y = "Vocabulary Words",
       color = "ID") +
  theme(legend.position = "bottom")
```



4.3 Model Building Adding Fixed Effects

Add in an a main effect for `sex` and store the result at `m1`.

```
m1 <- lmer(vocab ~ 1 + age12 + sex + (1 + age|pers),
           data = dat)
```

Now, let's interact `sex` with `age12`, stored as `m2`.

```
m2 <- lmer(vocab ~ 1 + age12*sex + (1 + age|pers),
           data = dat)
```

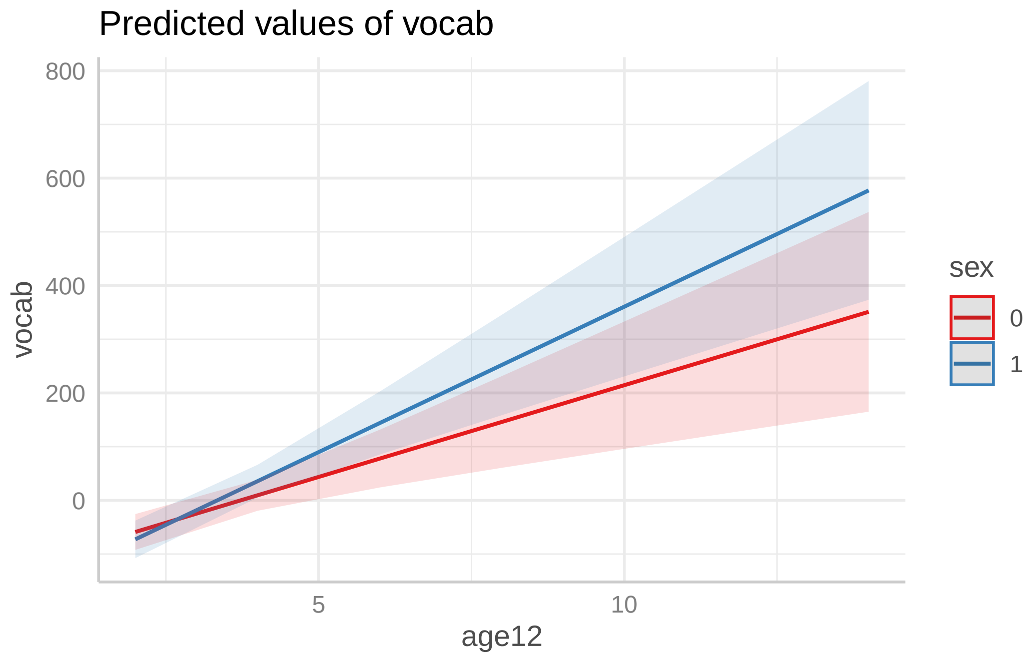
Use `screenreg` to tabulate the results.

```
screenreg(list(m1, m2))
##
## ======================================================
##                              Model 1       Model 2
## ------------------------------------------------------
## (Intercept)                 -156.58 ***   -127.21 ***
##                              (25.47)       (30.62)
## age12                         43.26 ***     34.16 ***
```

```
##                                (6.80)        (8.57)
## sex                            10.45        -53.71
##                               (18.56)       (44.93)
## age12:sex                                    19.99
##                                             (12.70)
## ------------------------------------------------------
## AIC                           759.67        752.38
## BIC                           775.51        770.48
## Log Likelihood               -372.83       -368.19
## Num. obs.                      71            71
## Num. groups: pers              11            11
## Var: pers (Intercept)      112163.47      96301.55
## Var: pers age                 493.78        424.88
## Cov: pers (Intercept) age   -7432.44      -6387.01
## Var: Residual                1346.90       1348.99
## ======================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

The code below graphs the interaction model. Let's practice connecting the graph to the output.

```
ggeffect(m2, terms = c("age12", "sex")) |>
  plot()
```



Predicted values of vocab
800
600
400
200
0
vocab
5
10
age12
sex
0
1