# Conditional Models

Emorie D Beck

9/7/2017

# Packages

```
library(psych)
library(sjPlot)
library(broom)
library(lme4)
library(MuMIn)
library(merTools)
library(reghelper)
library(stargazer)
library(lsmeans)
library(multcompView)
library(plyr)
library(tidyverse)
```

# Basic Syntex

From last week:

- **Level 1:** $Y_{ij} = \beta_{0j} + \varepsilon ij$
- **Level 2:** $\beta_{0j} = \gamma_{00} + U_{0j}$

# Sample Data

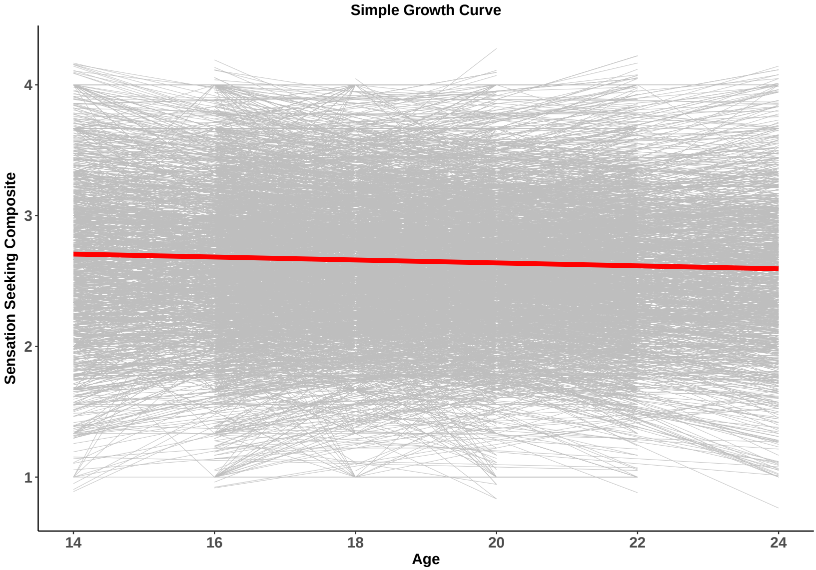
The National Longitudinal Study of Youths 1979 Child and Young Adult Sample (NLSYCYA) is a longitudinal study conducted by the National Bureau of Labor Statistics. The sample includes the children of the original 1979 sample. Here, we are going to use a subset of the more than 11,000 variables available that include the following.

| Item Name | Description | Time-Varying? |
|---|---|---|
| PROC_CID | Participant ID | No |
| Dem_DOB | Year of Date of Birth | No |
| groups | Jail, Community Service, None | No |
| DemPWeight | Weight Percentile at age 10 | No |
| age | Age of participant | Yes |
| Year | Year of Survey | Yes |
| age0 | Age of participant (centered) | Yes |
| SensSeek | Sensation-Seeking Composite | Yes |
| CESD | CESD Depression Composite | Yes |

# Simple Growth Curve Model

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * time_{ij} + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + U_{1j}$

# Simple Growth Curve Model



# In R

```r
mod0 <- lmer(SensSeek ~ age0 + (1|age0), data = sample_dat)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SensSeek ~ age0 + (1 | age0)
##    Data: sample_dat
##
## REML criterion at convergence: 28493.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.99678 -0.61465  0.01054  0.59619  2.47175
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  age0     (Intercept) 3.053e-15 5.525e-08
##  Residual             3.240e-01 5.692e-01
## Number of obs: 16645, groups:  age0, 6
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  2.705676   0.007061   383.2
## age0        -0.011252   0.001388    -8.1
##
## Correlation of Fixed Effects:
##      (Intr)
## age0 -0.781
```

# Conditional Models: Adding Predictors

Let's see if we can better predict participants' change in sensation seeking over time by adding covariates.

| Predictor | Continuous | Categorical |
|---|---|---|
| Time Invariant | Weight for Age | Group |
| Time Varying | CESD Scores | Depression |

# Time Invariant Predictors

# Time Invariant Predictors: Continuous

The basic equation, specifying a random intercept and slope:

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * time_{1j} + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} * X_{2j} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + \gamma_{11} * X_{2j} + U_{1j}$

But we need to break this down to see that adding additional predictors results in interaction terms:

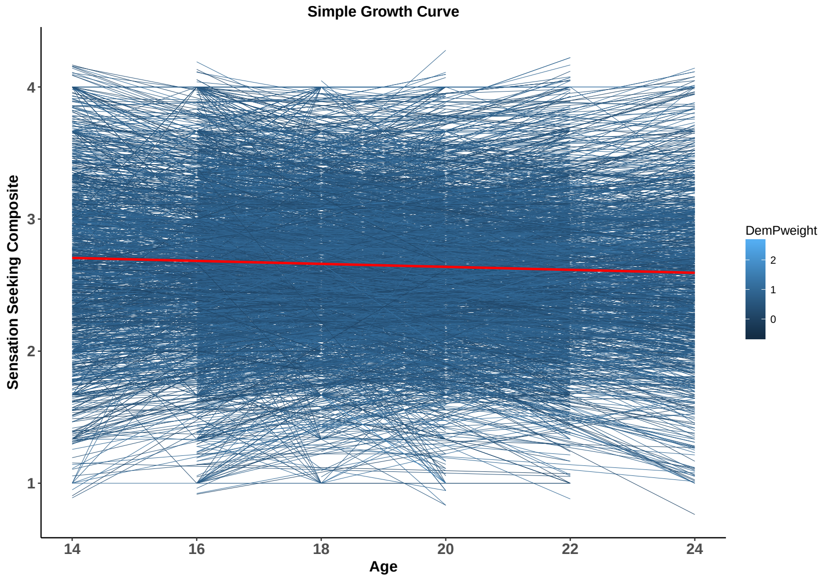
$$Y_{ij} = \gamma_{00} + \gamma_{01} * X_{2j} + U_{0j} + (\gamma_{10} + \gamma_{11} * X_{2j} + U_{1j}) * X_{1j} + \varepsilon ij$$
$$Y_{ij} = \gamma_{00} + \gamma_{01} * X_{2j} + \gamma_{10} * X_{1j} + \gamma_{11} * X_{2j} * X_{1j} + U_{0j} + U_{1j} * X_{1j} + \varepsilon ij$$

# Time Invariant Predictors: Continuous Example - Weight for Age Percentile

```r
describe(sample_dat$DemPweight)
```

```
##    vars     n mean   sd median trimmed  mad   min max range skew kur
## X1    1 16645 0.66 0.32   0.71    0.68 0.36 -0.64 2.7  3.35 -0.3
##    se
## X1  0
```

# Time Invariant Predictors: Continuous Example - Weight for Age Percentile



# Time Invariant Predictors: Continuous Example - Weight for Age Percentile

```r
# time invariant covariate with random intecept but not slope
mod1a <- lmer(SensSeek ~ age0 + DemPweight + age0*DemPweight + (1|PROC_CID),
              data = sample_dat)

# time invariant predictor with random slope and intercept
mod1b <- lmer(SensSeek ~ age0 + DemPweight + age0*DemPweight +
                (age0|PROC_CID), data = sample_dat)
```

# Time Invariant Predictors: Categorical Example - 2 level group

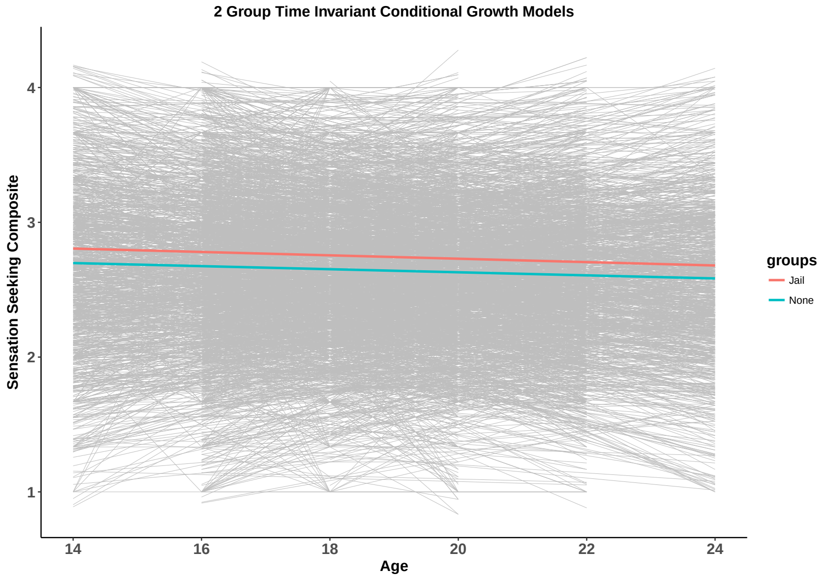
Lets's start with 2 groups: Jail v. None

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * time_{1j} + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} * X_{2j} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + \gamma_{11} * X_{2j} + U_{1j}$

# Time Invariant Predictors: Example - 2 level group

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * age0_{ij} + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} * groupsNone + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + \gamma_{11} * groupsNone + U_{1j}$

| Variable | D1 |
|---|---|
| Jail | 0 |
| None | 1 |



2 Group Time Invariant Conditional Growth Models
Sensation Seeking Composite
4
3
2
1
14
16
18
20
22
24
Age
groups
Jail
None

# Time Invariant Predictors: Example - 2 level group

```r
mod2g <- lmer(SensSeek ~ age0 + groups + age0*groups + (age0|PROC_CID),
              data = sample_dat %>% filter(groups != "CommServ"))
summary(mod2g)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SensSeek ~ age0 + groups + age0 * groups + (age0 | PROC_CID)
##    Data: sample_dat %>% filter(groups != "CommServ")
##
## REML criterion at convergence: 25439.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8386 -0.4713  0.0097  0.4741  3.5087
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  PROC_CID (Intercept) 0.1601936 0.40024
##           age0        0.0009029 0.03005  -0.40
##  Residual             0.1757529 0.41923
## Number of obs: 16134, groups:  PROC_CID, 7188
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)       2.811511   0.030945   90.86
## age0             -0.011601   0.005078   -2.28
## groupsNone       -0.115815   0.031847   -3.64
## age0:groupsNone   0.002546   0.005231    0.49
##
## Correlation of Fixed Effects:
##             (Intr) age0   grpsNn
## age0        -0.633
## groupsNone  -0.972  0.615
## age0:grpsNn  0.615 -0.971 -0.634
```

# Side Note: `lme4` helper functions

```
vcov(mod2g)
VarCorr(mod2g)
fixef(mod2g)
ranef(mod2g)
coef(mod2g)
confint.merMod(mod2g, method = "boot")
reghelper::ICC(mod2g)
MuMIn::r.squaredGLMM(mod2g)
```

```r
vcov(mod2g)
```

```
## 4 x 4 Matrix of class "dpoMatrix"
##                    (Intercept)          age0    groupsNone age
## (Intercept)       9.575941e-04 -9.950566e-05 -9.575941e-04
## age0             -9.950566e-05  2.578270e-05  9.950566e-05   -
## groupsNone       -9.575941e-04  9.950566e-05  1.014221e-03   -
## age0:groupsNone   9.950566e-05 -2.578270e-05 -1.055740e-04
```

```r
VarCorr(mod2g)
```

```
##  Groups   Name        Std.Dev. Corr
##  PROC_CID (Intercept) 0.400242
##           age0        0.030048 -0.397
##  Residual             0.419229
```

```r
fixef(mod2g)
```

```
##      (Intercept)             age0       groupsNone age0:groupsNo
##      2.811510889     -0.011600946     -0.115815331      0.0025459
```

```r
ranef(mod2g)
```

```
## $PROC_CID
##           (Intercept)          age0
## 301      0.2386857200 -2.795195e-03
## 302      0.0209145813  7.910316e-04
## 303     -0.1976874625  1.107953e-03
## 401      0.4385500788 -9.212628e-03
## 801      0.0521633381 -1.724546e-02
## 802      0.5713034994 -1.458977e-02
## 803      0.0540967467 -8.653343e-03
## 1001     0.3045440068 -2.515398e-03
## 1602     0.1451053206 -4.326995e-03
## 1603     0.1515629720 -2.989324e-03
## 2001    -0.1018502691 -5.327357e-05
## 2501    -0.2280747057  6.426858e-03
## 2502     0.1451053206 -4.326995e-03
## 2701     0.1223335127 -3.781162e-03
## 2702     0.1515629720 -2.989324e-03
## 3002    -0.0178411027 -1.821252e-02
## 3003    -0.2294312321  5.748447e-03
## 4301     0.3124375351  2.608497e-03
```

```r
confint.merMod(mod2g, method = "boot", nsim = 10)
```

```
##                         2.5 %       97.5 %
## .sig01             0.386164710  0.405363134
## .sig02            -0.421473723 -0.280943625
## .sig03             0.026226070  0.033223113
## .sigma             0.411804912  0.422225484
## (Intercept)        2.777500804  2.896510060
## age0              -0.020983416 -0.002090931
## groupsNone        -0.193656044 -0.073856220
## age0:groupsNone   -0.006936207  0.011103280
```

All units of the random effects are in standard deviation units (which means you need to square them to get the variance!!)

- .sig01 $=$ sd of random intercept $= \sqrt{\tau_{00}}$
- .sig02 $=$ correlation between slope and intercept $= \sqrt{\tau_{10}}$
- .sig03 $=$ sd of random slope $= \sqrt{\tau_{11}}$
- .sigma $=$ residual variance $= \hat{\sigma}$

```r
reghelper::ICC(mod2g)
```

```
## [1] 0.4782449
```

**Conditional** $R^2$**:** how well the model fits overall

**Marginal** $R^2$**:** how well the model fits the fixed effects

```
MuMIn::r.squaredGLMM(mod2g)
```

```
##          R2m         R2c
## 0.004487822 0.455541793
```

# Side Note: Creating MLM Tables

There are lots of helpful packages for this, including `stargazer` and `sjPlot`, which are demonstrated below.

```
stargazer::stargazer(mod2g)
sjPlot::sjt.lmer(mod2g)
```

The problem is that `stargazer()` doesn't include all the terms we want, and `sjt.lmer()` only renders html. Embedded in the `.Rmd` version of these slides is some code that should help you to extract the terms you need and create a table using `dplyr` and `tidyr` that you can render in LaTeXusing `stargazer`.

# Side Note: Creating MLM Tables

But let's understand where those variables came from. To do so, we'll use the `broom` package in R to grab the terms we need.

| Description | Math Notation |
|---|---|
| Fixed Effect Intercept | $\gamma_{00}$ |
| Fixed Effect Group Intercept | $\gamma_{01}$ |
| Fixed Effect Age Slope | $\gamma_{10}$ |
| Fixed Effect Group Slope | $\gamma_{11}$ |
| Individual Random Intercepts | $U_{0j}$ |
| Variance of Random Intercepts | $\tau_{00}$ |
| Random Age Slopes | $U_{10}$ |
| Variance of Random Age Slopes | $\tau_{11}$ |
| Correlation b/w Random Slopes and Intercepts | $\tau_{10}$ |
| Residual Variance | $\hat{\sigma}^2$ |
| Intraclass Correlation | ICC |
| Conditional $R^2$ | $R^2_c$ |
| Marginal $R^2$ | $R^2_m$ |

# Side Note: Creating MLM Tables

```r
broom::tidy(mod2g)
broom::glance(mod2g)
```

```
##                             term     estimate   std.error  statistic
## 1                    (Intercept)  2.811510889 0.030945018 90.8550409
## 2                           age0 -0.011600946 0.005077667 -2.2846999
## 3                     groupsNone -0.115815331 0.031846832 -3.6366359
## 4                age0:groupsNone  0.002545993 0.005231305  0.4866841
## 5        sd_(Intercept).PROC_CID  0.400241919          NA         NA
## 6               sd_age0.PROC_CID  0.030048359          NA         NA
## 7 cor_(Intercept).age0.PROC_CID -0.397195984          NA         NA
## 8       sd_Observation.Residual  0.419228923          NA         NA
##      group
## 1    fixed
## 2    fixed
## 3    fixed
## 4    fixed
## 5 PROC_CID
## 6 PROC_CID
## 7 PROC_CID
## 8 Residual
```

```
##       sigma    logLik      AIC      BIC deviance df.residual
## 1 0.4192289 -12719.82 25455.64 25517.15  25405.3       16126
```

# Side Note: Creating MLM Tables

```r
options(knitr.kable.NA = '')
knitr::kable(tab, caption = "Ugly MLM Table Example")
```

Table 5: Ugly MLM Table Example

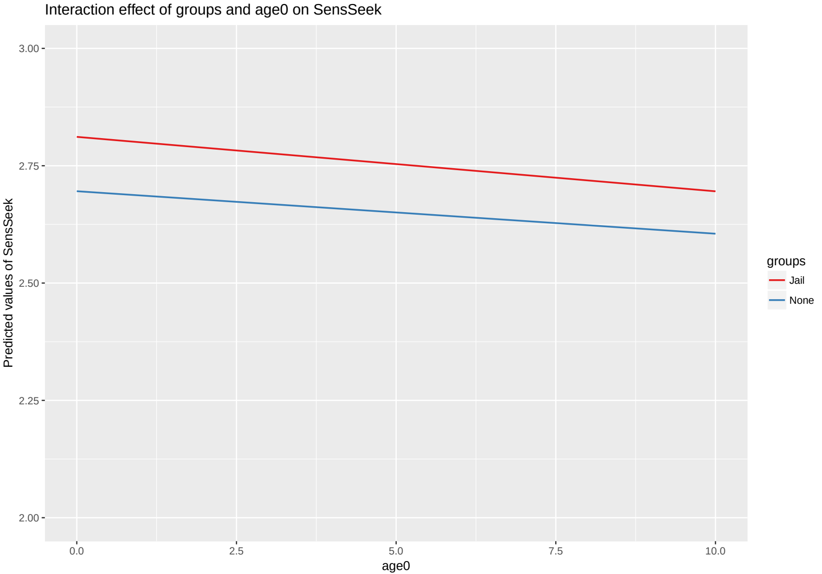
| type | term | estimate | CI |
|---|---|---|---|
| Fixed Parts | (Intercept) | 2.81 | (2.77, 2.86) |
| Fixed Parts | age0 | -0.01 | (-0.02, -0.00) |
| Fixed Parts | groupsNone | -0.12 | (-0.15, -0.07) |
| Fixed Parts | age0:groupsNone | 0.00 | (-0.01, 0.01) |
| Random Parts | $\tau$00 | 0.16 | (0.16, 0.17) |
| Random Parts | $\tau_{11}$ | 0.00 | (0.00, 0.00) |
| Random Parts | $\tau_{10}$ | 0.16 | (0.21, 0.09) |
| Random Parts | $\hat{\sigma}^2$ | 0.18 | (0.17, 0.18) |
| mod_terms | ICC | 0.48 | |
| mod_terms | $R_m^2$ | 0.00 | |
| mod_terms | $R_c^2$ | 0.46 | |

# Side Note: Plotting Simple Effects

```r
# categorical
sjp.int(mod2g, type = "eff", p.kr = F, swap.pred = T)
# continuous
sjp.int(mod1b, type = "eff", p.kr = F, swap.pred = T,
        mdrt.values = "meansd")
```

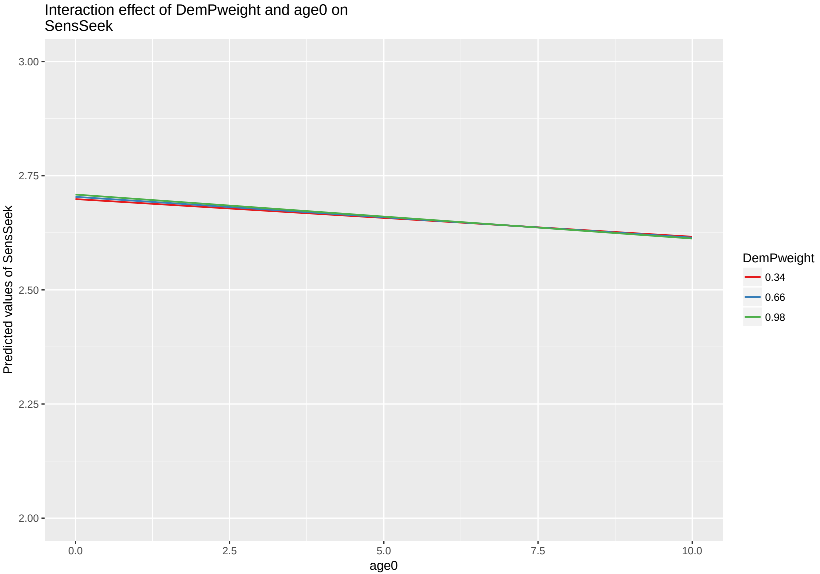
# Side Note: Plotting Simple Effects (Categorical)

```
sjp.int(mod2g, type = "eff", p.kr = F, swap.pred = T)
```



# Side Note: Plotting Simple Effects (Continuous)

```
sjp.int(mod1b, type = "eff", p.kr = F, swap.pred = T, mdrt.value
```



# Side Note: Comparisons with `lsmeans`

```r
# create a reference grid
ref.grid2g <- ref.grid(mod2g)
# create the lsmeans object
lsgroups   <- lsmeans(ref.grid2g, "groups")
# compact letter display
cld(lsgroups, alpha = .10)
# plot
plot(lsgroups)
# contrasts of the ref.grid object
contrast(ref.grid2g, method = "eff")
# comparisons
groups.sum <- summary(lsgroups, infer = c(TRUE,TRUE),
                      level = .90, adjust = "bon", by = "groups")
```

```r
# create a reference grid
(ref.grid2g <- ref.grid(mod2g))
```

```
## 'ref.grid' object with variables:
##     age0 = 3.9838
##     groups = Jail, None
```

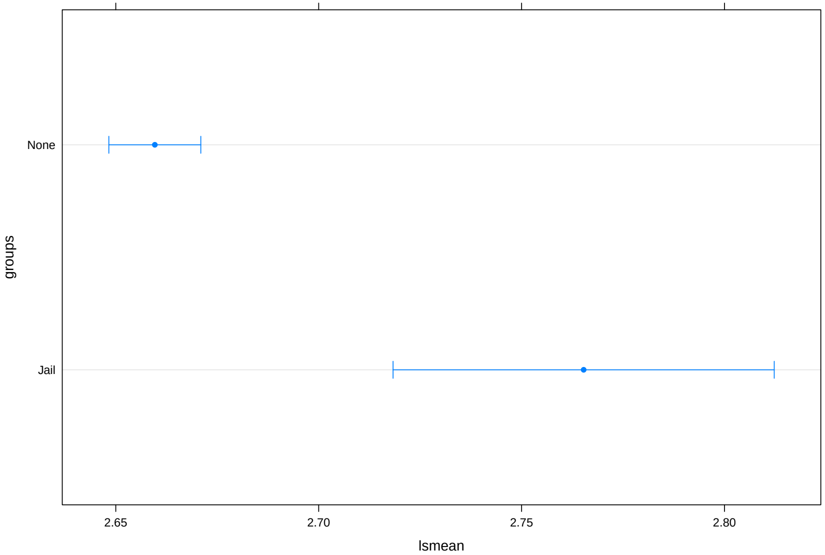
```r
# create the lsmeans object
(lsgroups   <- lsmeans(ref.grid2g, "groups"))
```

```
##  groups   lsmean          SE      df lower.CL upper.CL
##  Jail   2.765295 0.023957487 6923.38 2.718331 2.812260
##  None   2.659623 0.005780429 7125.64 2.648291 2.670954
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
```

```r
# compact letter display
cld(lsgroups, alpha = .10)
```

```
##  groups   lsmean          SE      df lower.CL upper.CL .
##  None   2.659623 0.005780429 7125.64 2.648291 2.670954
##  Jail   2.765295 0.023957487 6923.38 2.718331 2.812260
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
## significance level used: alpha = 0.1
```

```r
# plot
plot(lsgroups)
```



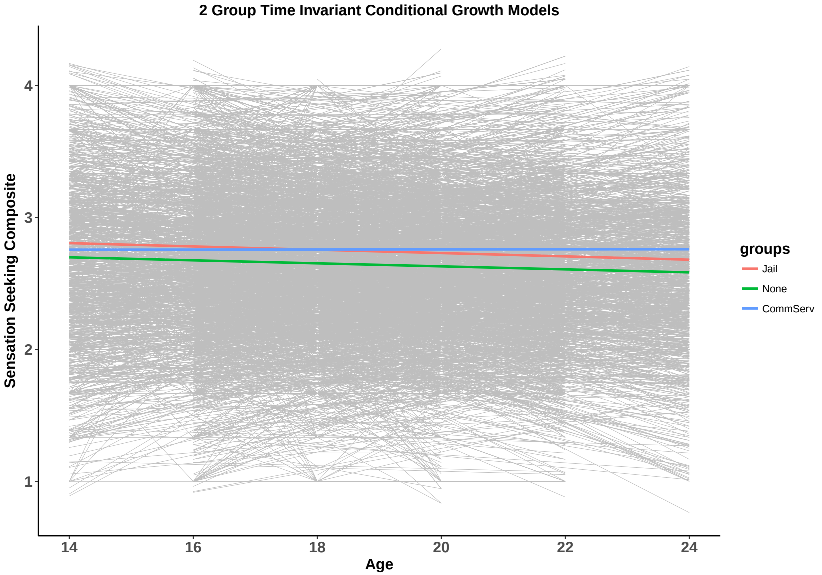
```r
# contrasts of the ref.grid object
contrast(ref.grid2g, method = "eff")
```

```
##  contrast                              estimate         SE
##  3.9837610016115,Jail effect  0.05283635 0.01232249 6934
##  3.9837610016115,None effect -0.05283635 0.01232249 6934
##  p.value
##   <.0001
##   <.0001
##
## P value adjustment: fdr method for 2 tests
```

```r
# comparisons
(groups.sum <- summary(lsgroups, infer = c(TRUE,TRUE),
          level = .90, adjust = "bon", by = "groups"))
```

```
## groups = Jail:
##    lsmean          SE      df lower.CL upper.CL t.ratio
##  2.765295 0.023957487 6923.38 2.725884 2.804707 115.425
##
## groups = None:
##    lsmean          SE      df lower.CL upper.CL t.ratio
##  2.659623 0.005780429 7125.64 2.650114 2.669132 460.108
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.9
```

# Time Invariant Predictors: Example - 3 level group



# Time Invariant Predictors: Example - 3 level group

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * age0_{ij} + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} * D1 + \gamma_{02} * D2 + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + \gamma_{11} * D1 + \gamma_{12} * D2 + U_{1j}$

| Variable | D1 | D2 |
|---|---|---|
| Jail | 0 | 0 |
| None | 1 | 0 |
| CommServ } 0 | 1 | |

# Time Invariant Predictors: Example - 3 level group

```
mod3g <- lmer(SensSeek ~ age0 + groups + age0*groups +
                (age0|PROC_CID), data = sample_dat)
summary(mod3g)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SensSeek ~ age0 + groups + age0 * groups + (age0 | PROC_CID)
##    Data: sample_dat
##
## REML criterion at convergence: 26251.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8387 -0.4727  0.0097  0.4728  3.5010
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  PROC_CID (Intercept) 0.1579466 0.39742
##           age0        0.0008989 0.02998  -0.39
##  Residual             0.1761279 0.41968
## Number of obs: 16645, groups:  PROC_CID, 7423
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)           2.811413   0.030865   91.09
## age0                 -0.011577   0.005082   -2.28
## groupsNone           -0.115728   0.031765   -3.64
## groupsCommServ       -0.058406   0.050348   -1.16
## age0:groupsNone       0.002527   0.005236    0.48
## age0:groupsCommServ   0.013322   0.008875    1.50
##
## Correlation of Fixed Effects:
##             (Intr) age0   grpsNn grpsCS ag0:gN
```

# Time Varying Predictors

# Time Varying Predictors: Continuous

Next, we'll add in a time-varying predictor. Maybe it's not that our participants sensation seeking is moderated by early life experiences of jail or court-ordered community service. Instead, their sensation seeking is moderated by depression.
How does this look?

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * time + \beta_{2j} * CESD + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + U_{1j}$
  - $\beta_{2j} = \gamma_{20}$

# Time Varying Predictors: Continuous

## To Interaction or Not - That Is the Question

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * age0 + \beta_{2j} * CESD + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + U_{1j}$
  - $\beta_{2j} = \gamma_{20}$

$$Y_{ij} = \gamma_{00} + \gamma_{01} + U_{0j} + (\gamma_{10} + U_{1j}) * age0 + \gamma_{20} * CESD$$

# Time Varying Predictors: Continuous

```r
modTV1 <- lmer(SensSeek ~ age0 + CESD + (age0|PROC_CID), data =
summary(modTV1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SensSeek ~ age0 + CESD + (age0 | PROC_CID)
##    Data: sample_dat
##
## REML criterion at convergence: 26222.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8913 -0.4785  0.0140  0.4705  3.5504
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  PROC_CID (Intercept) 0.1572108 0.39650
##           age0        0.0008911 0.02985  -0.38
##  Residual             0.1759002 0.41940
## Number of obs: 16645, groups:  PROC_CID, 7423
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  2.672365   0.008797  303.78
## age0        -0.009076   0.001205   -7.53
## CESD         0.049854   0.008088    6.16
##
## Correlation of Fixed Effects:
##      (Intr) age0
## age0 -0.507
## CESD -0.580 -0.020
```

# Time Varying Predictors: Categorical

Next, we'll add in a time-varying predictor. Maybe it's not that our participants sensation seeking is moderated by early life experiences of jail or court-ordered community service. Instead, their sensation seeking is moderated by depression.
How does this look?

- **Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j} * time + \beta_{2j} * depressed + \varepsilon ij$
- **Level 2:**
  - $\beta_{0j} = \gamma_{00} + \gamma_{01} + U_{0j}$
  - $\beta_{1j} = \gamma_{10} + U_{1j}$
  - $\beta_{2j} = \gamma_{20}$

# Time Varying Predictors: Categorical

```r
# creating a dummy variable for time varying categorical depress
sample_dat <- sample_dat %>%
  mutate(depressed =
           factor(ifelse(CESD <= 1.5, 0, 1), levels = c(0,1),
                  labels = c("Depressed", "Not Depressed")))
modTV2 <- lmer(SensSeek ~ age0 + depressed + (age0|PROC_CID),
               data = sample_dat)
summary(modTV2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SensSeek ~ age0 + depressed + (age0 | PROC_CID)
##    Data: sample_dat
##
## REML criterion at convergence: 26257.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8439 -0.4702  0.0091  0.4721  3.4959
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  PROC_CID (Intercept) 0.1583541 0.39794
##           age0        0.0008978 0.02996  -0.38
##  Residual             0.1761783 0.41974
## Number of obs: 16645, groups:  PROC_CID, 7423
##
## Fixed effects:
##                          Estimate Std. Error t value
```

# confidence intervals and effect size

# fitted / predicted values

# Other Things

# autoregressive models, autoregressive errors

# cohen's d - changing intercept with 0 at last time point