Assignment 3: ALDA

Abhilasha Kumar

September 27, 2017

1 Reading the Data

We first read a demographics file into our final data frame, so that we can use gender and age as covariates in our analysis.

```
> cell_demo = read.csv("cell_demo.csv", header = TRUE, sep = ",")
> cell = read.csv("cell_withitems_complete.csv", header = TRUE, sep = ",")
> cell = merge(cell, cell_demo, by = "ID")
> cell$ID = as.factor(as.character(cell$ID))
```

2 Time Invariant Nominal Covariate

We will use Gender as the time-invariant nominal covariate in this analysis. Our DV is TimeJudgment-Distance, and our IV is Days.

Predicting Only Intercept

```
> contrasts(cell$Gender) = contr.treatment(2)
> m1 = lmer(data = cell, TimeJudgmentDistance ~ Days + Gender + (1|ID))
> summary(m1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days + Gender + (1 | ID)
   Data: cell

REML criterion at convergence: 4001.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5085 -0.5722 -0.2381  0.1814  4.2472

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1792   0.4233
 Residual             6.3510   2.5201
Number of obs: 847, groups:  ID, 44

Fixed effects:
              Estimate Std. Error t value
(Intercept)  0.7522695  0.1975627   3.808
Days         0.0078621  0.0008239   9.543
Gender2     -0.2889115  0.2173066  -1.330

Correlation of Fixed Effects:
```

```
        (Intr) Days
Days    -0.648
Gender2 -0.541  0.020
```

Predicting Both Slope and Intercept

```
> m2 = lmer(data = cell, TimeJudgmentDistance ~ Days*Gender + (1|ID))
> summary(m2)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days * Gender + (1 | ID)
   Data: cell

REML criterion at convergence: 4007.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6376 -0.5620 -0.2496  0.1975  4.1198

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1805   0.4248
 Residual             6.3173   2.5134
Number of obs: 847, groups:  ID, 44

Fixed effects:
              Estimate Std. Error t value
(Intercept)   0.482226   0.229407   2.102
Days          0.009601   0.001115   8.611
Gender2       0.290974   0.332205   0.876
Days:Gender2 -0.003805   0.001650  -2.307

Correlation of Fixed Effects:
            (Intr) Days   Gendr2
Days        -0.755
Gender2     -0.691  0.521
Days:Gendr2  0.510 -0.676 -0.757
```

Centering

Since Gender is a nominal variable, it cannot be centered. We could potentially use effects coding instead of dummy coding, but that would not change the overall fit of the model itself, although it. Below, we effects code the Gender variable to see if it affects the model:

```
> contrasts(cell$Gender) = contr.sum(2)
> m2_effects = lmer(data = cell, TimeJudgmentDistance ~ Days*Gender + (1|ID))
> summary(m2_effects)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days * Gender + (1 | ID)
   Data: cell

REML criterion at convergence: 4010
```

```
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6376 -0.5620 -0.2496  0.1975  4.1198

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1805   0.4248
 Residual             6.3173   2.5134
Number of obs: 847, groups:  ID, 44

Fixed effects:
               Estimate Std. Error t value
(Intercept)   0.6277132  0.1661026   3.779
Days          0.0076985  0.0008248   9.334
Gender1      -0.1454868  0.1661026  -0.876
Days:Gender1  0.0019027  0.0008248   2.307

Correlation of Fixed Effects:
            (Intr) Days   Gendr1
Days        -0.757
Gender1     -0.046  0.052
Days:Gendr1  0.052 -0.086 -0.757
```

Notice that the interpretation of the coefficients changes, in that earlier the intercept was the value for the dummy coded Gender variable (0 : female), but now the intercept is the value of TimeJudgmentDistance at an average value of gender. Similarly, the Days coefficient is the increase in TimeJudgmentDistance, for an average value of gender.

To know whether the model with intercept-only or both slope and intercept fits the data better, we run an ANOVA:

```
> anova(m1,m2)
```

```
Data: cell
Models:
m1: TimeJudgmentDistance ~ Days + Gender + (1 | ID)
m2: TimeJudgmentDistance ~ Days * Gender + (1 | ID)
   Df    AIC    BIC  logLik deviance  Chisq Chi Df
m1  5 3995.3 4019.0 -1992.6   3985.3
m2  6 3992.0 4020.4 -1990.0   3980.0 5.3202      1
   Pr(>Chisq)
m1
m2    0.02108 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Thus, the model in which Gender predicts both slope and intercept explains more of the variance, and it is thus our final model.

3 Time Invariant Continuous Covariate

Predicting Only Intercept

```
> m3 = lmer(data = cell, TimeJudgmentDistance ~ Days + Age + (1|ID))
> summary(m3)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days + Age + (1 | ID)
   Data: cell

REML criterion at convergence: 4008.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4905 -0.5681 -0.2490  0.1893  4.2111

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1955   0.4422
 Residual             6.3502   2.5200
Number of obs: 847, groups:  ID, 44

Fixed effects:
              Estimate Std. Error t value
(Intercept)  0.8613109  0.3820591   2.254
Days         0.0078423  0.0008263   9.491
Age         -0.0089452  0.0122275  -0.732

Correlation of Fixed Effects:
     (Intr) Days
Days -0.396
Age  -0.899  0.075
```

Predicting Both Slope and Intercept

```
> m4 = lmer(data = cell, TimeJudgmentDistance ~ Days*Age + (1|ID))
> summary(m4)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days * Age + (1 | ID)
   Data: cell

REML criterion at convergence: 4025

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4967 -0.5634 -0.2528  0.1902  4.2157

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1969   0.4437
 Residual             6.3561   2.5211
Number of obs: 847, groups:  ID, 44

Fixed effects:
              Estimate Std. Error t value
(Intercept)  0.7382823  0.5166217   1.429
Days         0.0087744  0.0027553   3.185
Age         -0.0044077  0.0177178  -0.249
Days:Age    -0.0000351  0.0000990  -0.355
```

```
Correlation of Fixed Effects:
         (Intr) Days   Age
Days     -0.729
Age      -0.946  0.705
Days:Age  0.672 -0.954 -0.723
fit warnings:
Some predictor variables are on very different scales: consider rescaling
```

We compare Models 3 and 4 to see whether the interaction term explains any more of the variance:

```
> anova(m3, m4)
```

```
Data: cell
Models:
m3: TimeJudgmentDistance ~ Days + Age + (1 | ID)
m4: TimeJudgmentDistance ~ Days * Age + (1 | ID)
   Df    AIC    BIC  logLik deviance  Chisq Chi Df
m3  5 3996.6 4020.3 -1993.3   3986.6
m4  6 3998.4 4026.9 -1993.2   3986.4 0.1217      1
   Pr(>Chisq)
m3
m4     0.7272
```

Since Model 4 is not significantly different from Model 3, we will pick Model 3 as our final model.

Centering

Now, we center the Age and Days variable.

```
> cell$Age.c = scale(cell$Age, center = TRUE, scale = FALSE)
> cell$Days.c = scale(cell$Days, center = TRUE, scale = FALSE)
> m5 = lmer(data = cell, TimeJudgmentDistance ~ Days.c*Age.c + (1|ID))
> summary(m5)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TimeJudgmentDistance ~ Days.c * Age.c + (1 | ID)
   Data: cell

REML criterion at convergence: 4025

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4967 -0.5634 -0.2528  0.1902  4.2157

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1969   0.4437
 Residual             6.3561   2.5211
Number of obs: 847, groups:  ID, 44

Fixed effects:
               Estimate Std. Error t value
(Intercept)   1.8247392  0.1108566  16.460
Days.c        0.0078271  0.0008279   9.455
Age.c        -0.0098129  0.0124857  -0.786
```

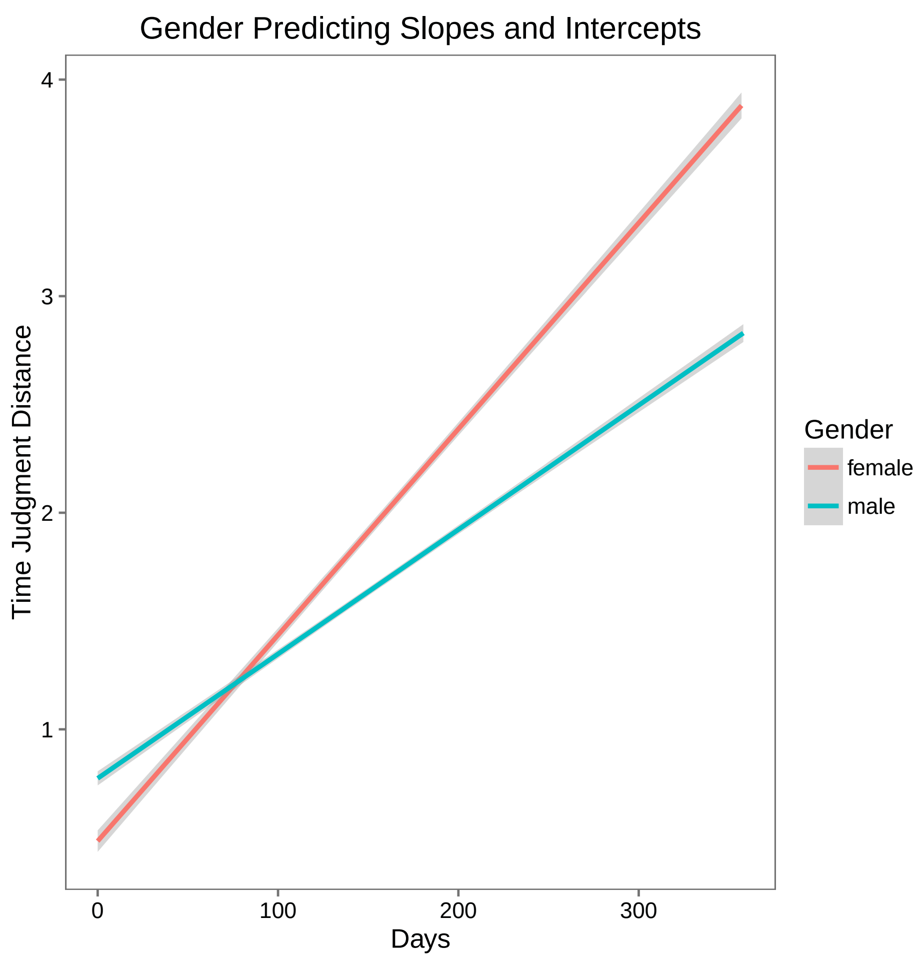
```
Days.c:Age.c -0.0000351  0.0000990  -0.355

Correlation of Fixed Effects:
            (Intr) Days.c Age.c
Days.c       0.011
Age.c       -0.020  0.084
Days.c:Ag.c  0.075  0.053  0.195
```

4 Graphing the Final Models

Nominal Covariate

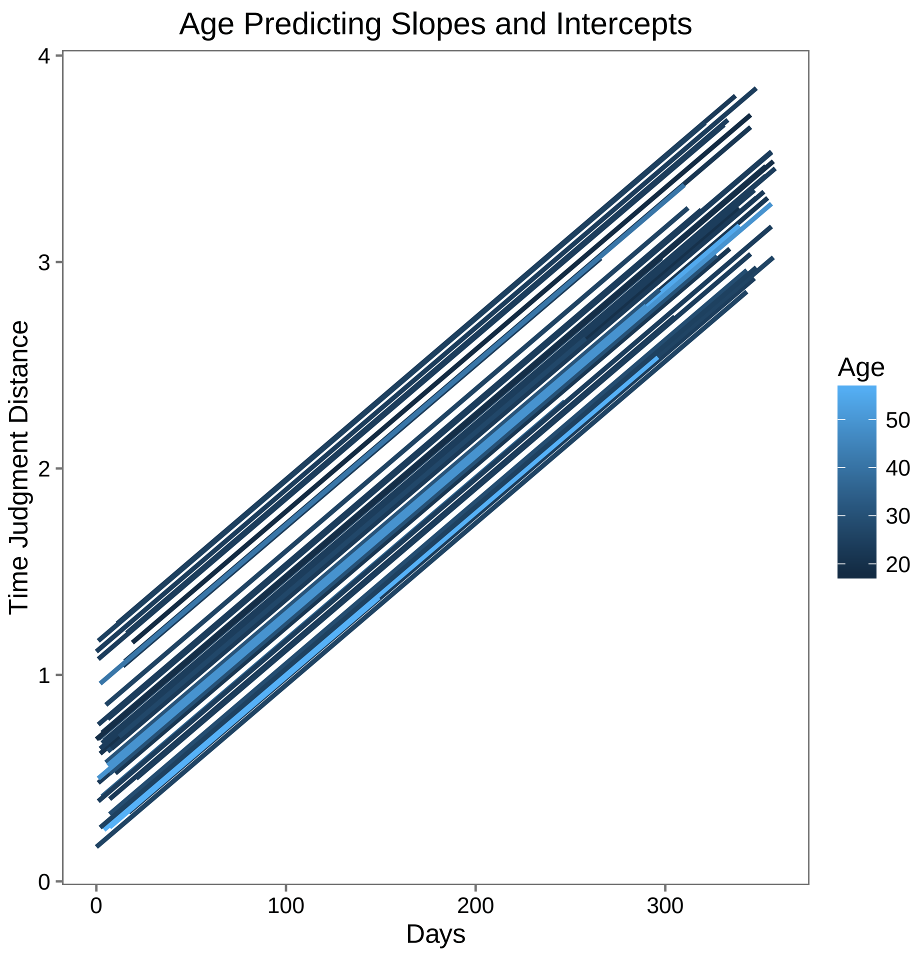
```
> library(broom)
> cell_nominal = augment(m2, cell)
> ggplot(cell_nominal, aes(x = Days, y = .fitted)) +
+   geom_smooth(method = "lm", aes(color = Gender)) +
+   xlab("Days") + ylab("Time Judgment Distance") +
+   theme_few()+
+   ggtitle("Gender Predicting Slopes and Intercepts")
```



Continuous Covariate

```
> library(broom)
> cell_categorical = augment(m3, cell)
```

```
> ggplot(cell_categorical, aes(x = Days, y = .fitted)) +
+   geom_smooth(method = "lm", aes(group = ID, color = Age)) +
+   xlab("Days") + ylab("Time Judgment Distance") +
+   theme_few()+
+   ggtitle("Age Predicting Slopes and Intercepts")
```



5 Calculating Confidence Intervals

We have Model 2 for the nominal covariate, and Model 3 for the categorical covariate. To calculate confidence intervals for our effects, we use the following equation:

$$\gamma_{00} \pm 1.96 * (\tau_{U_{0j}}) \tag{1}$$

Nominal Covariate

First for the intercept:

```
> 0.48226 + (1.96*0.4248)
```

```
[1] 1.314868
```

```
> 0.48226 - (1.96*0.4248)
```

```
[1] -0.350348
```

Then the slope:

```
> ## For days
> 0.009601 + (1.96*0.4248)
```

```
[1] 0.842209
```

```
> 0.009601 - (1.96*0.4248)
```

```
[1] -0.823007
```

```
> ## For gender
>
> 0.290974 + (1.96*0.4248)
```

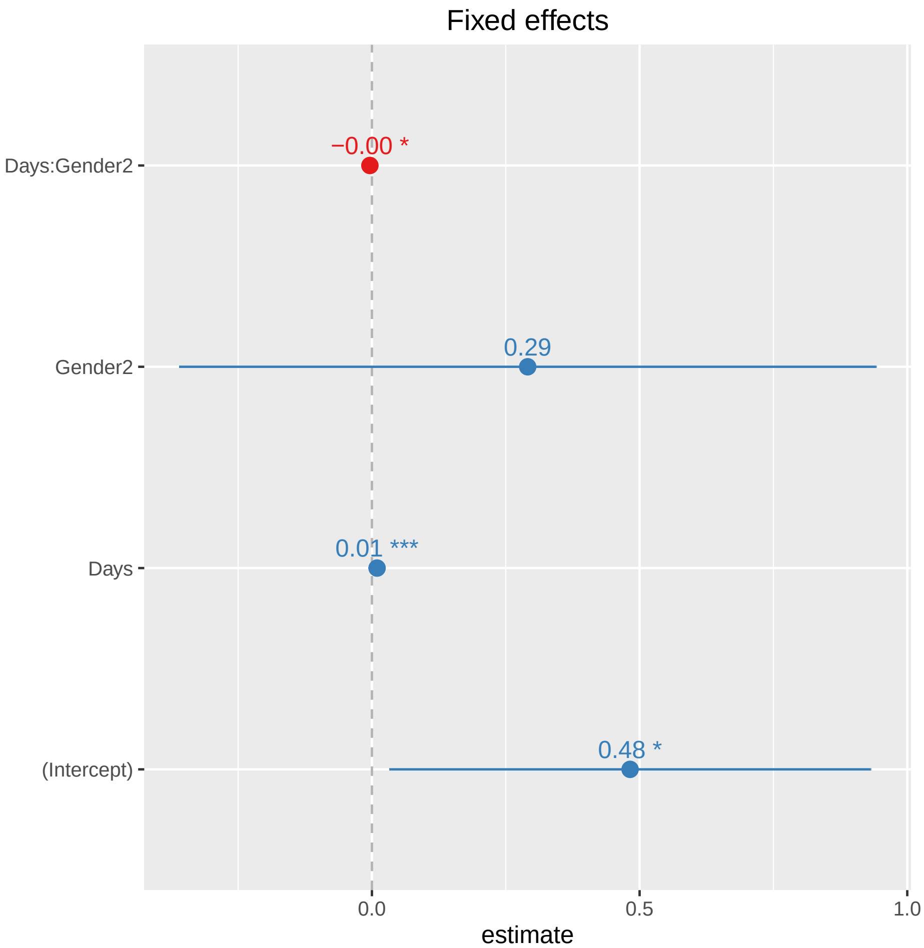
```
[1] 1.123582
```

```
> 0.290974 - (1.96*0.4248)
```

```
[1] -0.541634
```

Another way to visualize these confidence intervals is using the sjPlot function:

```
> sjp.lmer(m2, type = "fe")
```



Continuous Covariate

First for the intercept:

```
> 0.8613109 + (1.96*0.4422)
```

```
[1] 1.728023
```

```
> 0.8613109 - (1.96*0.4422)
```

```
[1] -0.0054011
```

Then for the slope:

```
> ## For days:
> 0.0078423 + (1.96*0.4422)
```

```
[1] 0.8745543
```

```
> 0.0078423 - (1.96*0.4422)
```

```
[1] -0.8588697
```

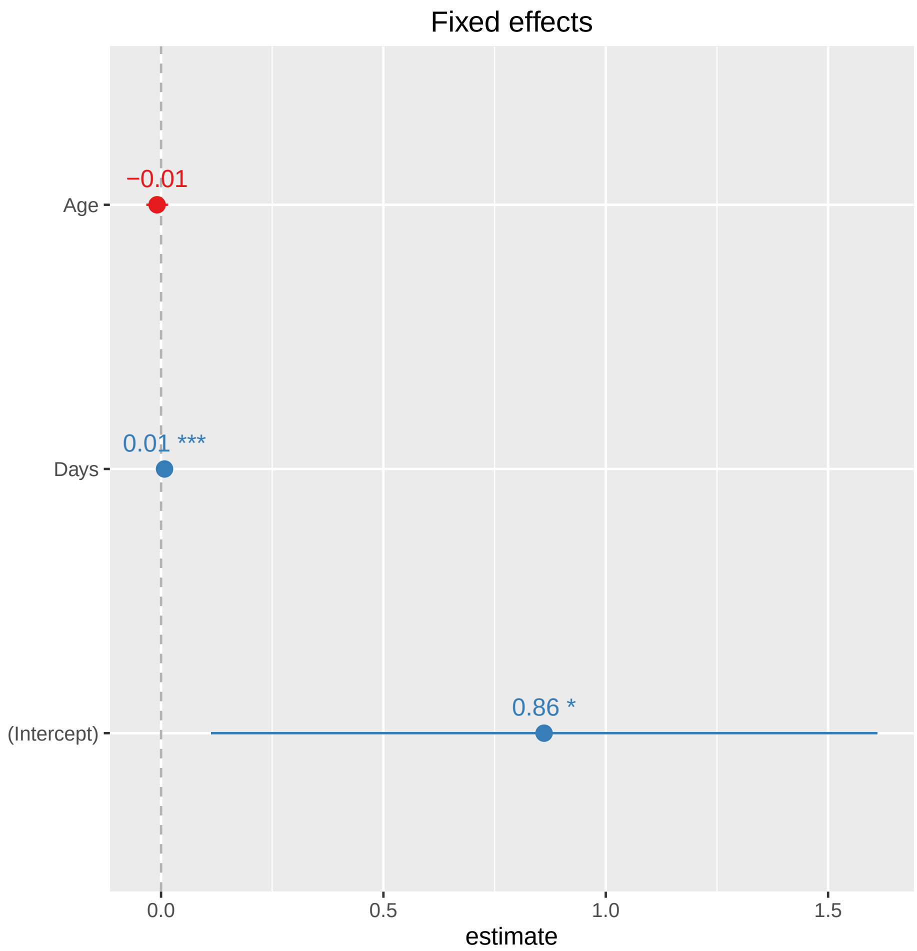
```
> ## For age
> -0.0089452 + (1.96*0.4422)
```

```
[1] 0.8577668
```

```
> -0.0089452 - (1.96*0.4422)
```

```
[1] -0.8756572
```

```
> sjp.lmer(m3, type = "fe")
```



6 Both Covariates

```
> m6 = lmer(data = cell, TimeJudgmentDistance ~ Days*Gender + Age + (1|ID))
> summary(m6)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula:
TimeJudgmentDistance ~ Days * Gender + Age + (1 | ID)
   Data: cell

REML criterion at convergence: 4016.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6450 -0.5579 -0.2396  0.2014  4.1103

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1876   0.4332
 Residual             6.3167   2.5133
Number of obs: 847, groups:  ID, 44

Fixed effects:
               Estimate Std. Error t value
(Intercept)   0.8703844  0.3799140   2.291
Days          0.0076556  0.0008273   9.254
Gender1      -0.1512560  0.1667972  -0.907
Age          -0.0086330  0.0121454  -0.711
Days:Gender1  0.0019138  0.0008251   2.320

Correlation of Fixed Effects:
            (Intr) Days   Gendr1 Age
Days        -0.398
Gender1     -0.064  0.055
Age         -0.899  0.076  0.048
Days:Gendr1  0.039 -0.087 -0.754 -0.019
```

The interpretation of parameters when both covariates are in the model changes, because the coefficients are now at constant values of the other covariate. For example, in Model 6:

The intercept is the value of TimeJudgmentDistance for females, at age = 0. Note that is this meaningless at this point, because our age variable is not centered.

Similarly, the Gender2 coefficient is the difference in TimeJudgmentDistance between males and females at age=0.

The Age coefficient is the decrease in TimeJudgmentDistance for every 1-unit increase in Age, for females (dummy coded 0)

The interaction term denotes the difference in slopes between males and females, at age = 0.

7 Time Varying Covariate

Suppose we wanted to covary out the number of messages sent to the person – this is a time-varying covariate in our data.

```
> cell$Messages.c = scale(cell$Messages, center = TRUE, scale = FALSE)
> m7 = lmer(data = cell, TimeJudgmentDistance ~ Days.c*Messages.c + (1|ID))
> summary(m7)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula:
TimeJudgmentDistance ~ Days.c * Messages.c + (1 | ID)
   Data: cell
```

```
REML criterion at convergence: 4028.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4837 -0.5795 -0.2530  0.1895  4.2133

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.1908   0.4368
 Residual             6.3648   2.5229
Number of obs: 847, groups:  ID, 44

Fixed effects:
                    Estimate Std. Error t value
(Intercept)        1.822e+00  1.189e-01  15.319
Days.c             7.855e-03  9.070e-04   8.660
Messages.c        -8.363e-04  1.047e-02  -0.080
Days.c:Messages.c -4.804e-06  7.889e-05  -0.061

Correlation of Fixed Effects:
            (Intr) Days.c Mssgs.
Days.c      0.143
Messages.c  0.354  0.410
Dys.c:Mssg. 0.383  0.354  0.926
fit warnings:
Some predictor variables are on very different scales: consider rescaling
```

We still get a rescaling warning, and so we may want to consider scaling i.e. z-scoring the variables. That might make the most sense for these models.