

# North Carolina Data

STAT 222 week 2

```
> #NC comparing models
> week3NC = read.table(file="http://www-stat.stanford.edu/~rag/stat222/ncLong.dat", header = T) long form
> week3NC$timeInt = week3NC$time - 1
> head(week3NC)
  ID time  Y  Z timeInt
1 705810  1 380 120     0
2 705810  2 377 120     1
3 705810  3 460 120     2
4 705810  4 472 120     3
5 705810  5 495 120     4
6 705810  6 566 120     5
> summary(week3NC)
  ID           time           Y           Z           timeInt
Min.   : 705810   Min.   :1.00   Min.   :270.0   Min.   : 64.0   Min.   :0.00
1st Qu.: 847813   1st Qu.:2.75   1st Qu.:395.0   1st Qu.: 97.0   1st Qu.:1.75
Median :1046817   Median :4.50   Median :464.0   Median :106.0   Median :3.50
Mean   :1461655   Mean   :4.50   Mean   :469.9   Mean   :106.1   Mean   :3.50
3rd Qu.:1290819   3rd Qu.:6.25   3rd Qu.:540.0   3rd Qu.:115.0   3rd Qu.:5.25
Max.   :11090821   Max.   :8.00   Max.   :762.0   Max.   :145.0   Max.   :7.00
```

malse initial time => (implicit intercept)

Descriptives in posted version

```
> attach(week3NC)
> xtabs(~ ID + timeInt, week3NC) # balanced, complete data, Bates does this way
```

see plots

# week 2 ~~descriptives~~ descriptives, SFYS analyses; here revisit the mixed-models

balanced data

```
> library(lme4)
> ncUnc = lmer(Y ~ timeInt + (1 + timeInt | ID), data = week3NC)
> summary(ncUnc)
Linear mixed model fit by REML
Formula: Y ~ timeInt + (1 + timeInt | ID)
Data: week3NC
AIC BIC logLik deviance REMLdev
20690 20724 -10339 20680 20678
Random effects:
Groups Name Variance Std.Dev. Corr
ID (Intercept) 326.059 18.0571
timeInt 46.229 6.7992 0.651
Residual 403.487 20.0870
```

no level 2

$$\hat{var}(0) 46.229 = 55.836 - MSR/SSR$$

see below

```
Number of obs: 2216, groups: ID, 277
> # best to transform time to be zero at the time point of interest
##### (cor(rate,initial status) correct, .651)
Fixed effects:
Estimate Std. Error t value
(Intercept) 342.300 1.336 256.27
timeInt 36.448 0.449 81.18
```

gets this correct m/c of  $cor(0, N(0))$  <sup>rate</sup> <sub>initial time</sub>

match SFYS

```
Correlation of Fixed Effects:
(Intr)
timeInt 0.279  $\hat{r}_{Int, rate}$  from lmlist
```

```
> ncCon = lmer(Y ~ timeInt + Z:timeInt + (1 + timeInt | ID), data = week3NC) # incl Z in slope L2
> summary(ncCon)
```

```
Linear mixed model fit by REML
Formula: Y ~ timeInt + Z:timeInt + (1 + timeInt | ID)
Data: week3NC
AIC BIC logLik deviance REMLdev
20582 20622 -10284 20565 20568
Random effects:
Groups Name Variance Std.Dev. Corr
ID (Intercept) 326.062 18.0572
timeInt 24.683 4.9681 0.262
Residual 403.486 20.0870
Number of obs: 2216, groups: ID, 277
Fixed effects:
Estimate Std. Error t value
(Intercept) 342.29994 1.33568 256.27
timeInt -0.95502 2.70714 -0.35
timeInt:Z 0.35266 0.02531 13.93
```

```
Correlation of Fixed Effects:
(Intr) timInt
timeInt -0.010
timeInt:Z 0.000 -0.992
```

```
> ncList = lmlist(Y ~ timeInt | ID, data = week3NC) # fit
> rate = coef(ncList)[2]
> var(rate)
timeInt
timeInt 55.83613
> ncList
Degrees of freedom: 2216 total; 1662 residual
Residual standard error: 20.08697
> sst = 2*(.5^2 + 1.5^2 + 2.5^2 + 3.5^2)
> sst
[1] 42
> 55.836 - (20.087)^2/42
[1] 46.22915
```

Level 2  $\alpha_0 = \gamma_{00} + u_0$   
 $\alpha_1 = \gamma_{10} + \gamma_{11}Z + u_1$

see rate / 2 plot

page 2 week 2

STAT222

North Carolina

```
> ncCon2 = lmer(Y ~ timeInt + Z*timeInt + (1 + timeInt | ID), data = week3NC) # incl Z in level,slope L2
```

```
> summary(ncCon2)
```

Linear mixed model fit by REML

```
Formula: Y ~ timeInt + Z * timeInt + (1 + timeInt | ID)
```

Data: week3NC

AIC BIC logLik deviance REMLdev

20501 20547 -10243 20478 20485

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
ID	(Intercept)	194.731	13.9546	
	timeInt	24.628	4.9626	0.379

Residual 403.486 20.0870

Number of obs: 2216, groups: ID, 277

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	254.32454	8.83286	28.793
timeInt	0.84389	2.71301	0.311
Z	0.82948	0.08258	10.045
timeInt:Z	0.33569	0.02536	13.235

Correlation of Fixed Effects:

	(Intr)	timInt	Z
timeInt	-0.067		
Z	-0.992	0.066	
timeInt:Z	0.066	-0.992	-0.067

```
> confint(ncUnc) #add-ons needed to do this
Error: $ operator not defined for this S4 class
```

new version works

```
> anova(ncUnc, ncCon, ncCon2) #formal model comparisons, nested trio
```

Data: week3NC

Models:

```
ncUnc: Y ~ timeInt + (1 + timeInt | ID)
```

```
ncCon: Y ~ timeInt + Z:timeInt + (1 + timeInt | ID)
```

```
ncCon2: Y ~ timeInt + Z * timeInt + (1 + timeInt | ID)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
ncUnc	6	20692	20727	-10340				
ncCon	7	20579	20619	-10282	115.33	1	< 2.2e-16	***
ncCon2	8	20494	20540	-10239	86.57	1	< 2.2e-16	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
> # ncCon2 seems a winner
```

```
> anova(ncCon, ncCon2) # just to show trio works as you would hope
```

Data: week3NC

Models:

```
ncCon: Y ~ timeInt + Z:timeInt + (1 + timeInt | ID)
```

```
ncCon2: Y ~ timeInt + Z * timeInt + (1 + timeInt | ID)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
ncCon	7	20579	20619	-10282				
ncCon2	8	20494	20540	-10239	86.57	1	< 2.2e-16	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
>
```

Level 1

$$y = \alpha_0 + \alpha_1 t + \epsilon$$

Level 2

$$\alpha_0 = \gamma_{00} + \gamma_{01} z + u_0$$

$$\alpha_1 = \gamma_{10} + \gamma_{11} z + u_1$$

Z w/ slope

vs

Z w/ int and slope