Looking at Longitudinal Data

Linear mixed-effects model and Multiple Imputation

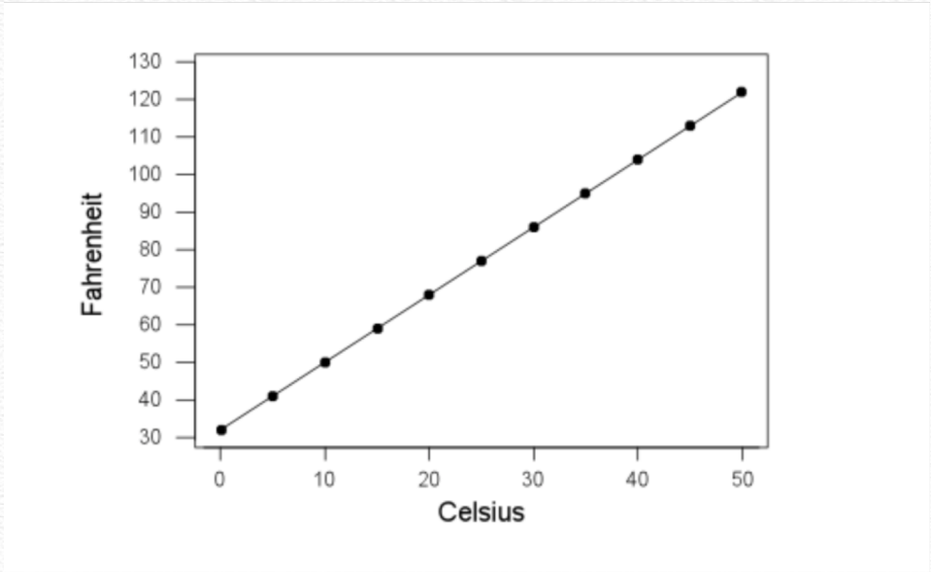
Monica

What is Longitudinal Data?

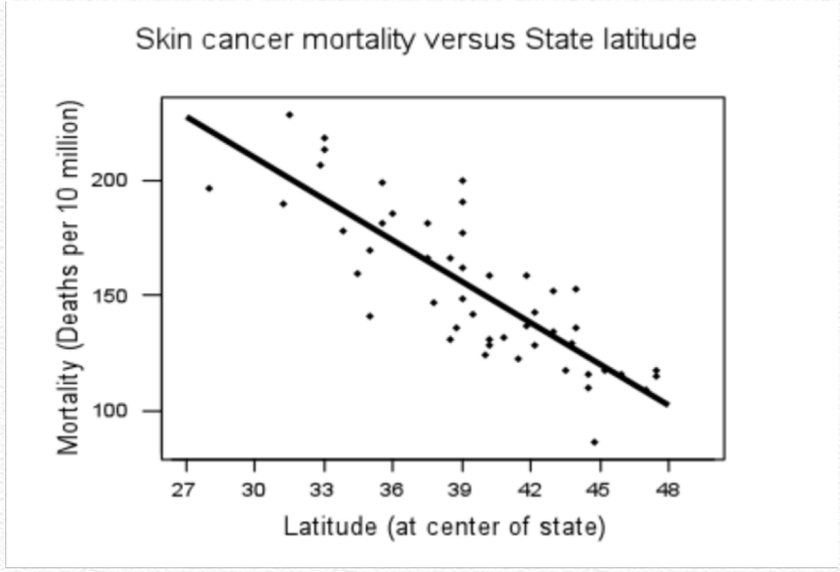
- Collection of repeated observations of the same sample at different time points.
- Design to measure intra-individual change over time
  - Lifespan issues
  - Treatment responses

General Linear Model

- Mathematical models that can describe functional and statistical relationship between variables.
  - Individual's Score on Outcome Measure = Overall constant + Effect for Measured Variable + Individual Error



$$F=\frac{9}{5}C+32$$



General Linear Model

- Multiple Linear Regression Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon.$$

- GLM Distributional Assumptions

$\varepsilon \sim N(0, \sigma^2)_{\ i.i.d}$

Linear mixed-effects model

- Combination of fixed and random effects

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

- Accounting for “subject” as a factor in the model to solve the problem

Multicollinearity: Variance Inflation Factors

- Correlations between predictor variables
  - Does not reduce the predictive power of the model as a whole
  - Affects calculation regarding individual predictors
  - Leads to less stable model
- VIF provides an index that measures the severity of collinearity

$$\mathrm{VIF_i} = \frac{1}{1 - R_i^2}$$

- `library(usdm)`

```
> vif(TwoBack[c("MADRS.Day", "Median.RTCR", "fc")])
    Variables      VIF
1   MADRS.Day 1.001830
2 Median.RTCR 1.065955
3          fc 1.067730
```

Centering and Scaling

- Centering: subtracting the mean of the variable from each data point so that the new variable's mean is 0
  - Makes interpretation of intercept easier
- Scaling: dividing each data point by a constant in order to alter the range of the data
  - Improve stability of model
  - Improve accuracy of parameter estimates

Centering and Scaling

```
dem = c("SFNB2.SFNB_RTC2", "fc")
demo[dem] <- lapply(demo[dem],scale)
```

SFNB2.SFNB_RTC2	fc
369.0000	0.128706
556.0000	0.710212
429.0000	0.433631
557.0000	0.301547
581.0000	0.271564
470.0000	0.425464
410.0000	0.662953
653.0000	0.504097
536.0000	0.067170
490.0000	0.272132
536.5000	0.758741
478.5000	0.799653

SFNB2.SFNB_RTC2	fc
-1.5049367	-1.11241806
0.3513656	1.10588339
-0.9093317	0.05079534
0.3612924	-0.45307243
0.5995344	-0.56745015
-0.5023349	0.01964025
-1.0979400	0.92560200
1.3142604	0.31960568
0.1528306	-1.34716267
-0.3037999	-0.56528337
0.1577940	1.29100951
-0.4179576	1.44707867

- Schielzeth, (2010)

Fitting Linear mixed-effects model in R

- Structure data into long format

	Record ID	Event..Name	Period	Site	Condition	MADRS.Day	MADRS.Raw.Score
1	P109	Baseline Visit	Period 1	fMRI Guided	Task	0	21
2	P109	Treatment Visit 2	Period 1	fMRI Guided	Task	1	21
3	P109	Treatment Visit 4	Period 1	fMRI Guided	Task	3	13
4	P109	Treatment Visit 6	Period 1	fMRI Guided	Task	7	12
5	P109	Treatment Visit 8	Period 1	fMRI Guided	Task	9	5
6	P109	Treatment Visit 10	Period 1	fMRI Guided	Task	13	8
7	P109	Phone Call Check-In Week 1	Period 1	fMRI Guided	Task	20	9
8	P109	Phone Call Check in Week 2	Period 1	fMRI Guided	Task	29	8
9	P109	Treatment Visit 12	Period 2	Standard (6cm)	Video	1	17
10	P109	Treatment Visit 14	Period 2	Standard (6cm)	Video	4	13
11	P109	Treatment Visit 16	Period 2	Standard (6cm)	Video	8	3
12	P109	Treatment Visit 18	Period 2	Standard (6cm)	Video	10	1
13	P109	Treatment Visit 20	Period 2	Standard (6cm)	Video	14	1
14	P109	Follow Up Visit 1	Period 2	Standard (6cm)	Video	53	0
15	P115	Baseline Visit	Period 1	fMRI Guided	Video	0	36
16	P115	Treatment Visit 2	Period 1	fMRI Guided	Video	12	30
17	P115	Treatment Visit 4	Period 1	fMRI Guided	Video	14	20
18	P115	Treatment Visit 6	Period 1	fMRI Guided	Video	18	6
19	P115	Treatment Visit 8	Period 1	fMRI Guided	Video	20	8
20	P115	Treatment Visit 10	Period 1	fMRI Guided	Video	22	5
21	P115	Phone Call Check-In Week 1	Period 1	fMRI Guided	Video	33	3
22	P115	Phone Call Check in Week 2	Period 1	fMRI Guided	Video	40	0

Fitting Linear mixed-effects model in R

- `library(lme4)`

- `library(lmerTest)`

```
> AllData.test1 = lmer(MADRS.Raw.Score ~ Site + Condition + MADRS.Day + (1|Record.ID), data=TwoBack,
REML=FALSE)
> summary(AllData.test1)
```

```
Random effects:
 Groups    Name        Variance Std.Dev.
 Record.ID (Intercept) 58.08    7.621
 Residual              22.58    4.752
Number of obs: 122, groups:  Record.ID, 25

Fixed effects:
                   Estimate Std. Error      df t value Pr(>|t|)
(Intercept)         23.2897     3.0078 27.0489   7.743 2.48e-08 ***
SiteStandard (6cm)   1.4813     3.1745 25.0437   0.467    0.645
ConditionVideo      -3.5384     3.1943 25.0219  -1.108    0.279
MADRS.Day           -0.6763     0.1053 97.3423  -6.420 4.98e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Comparing Models: Likelihood Ratio Test

- Goodness-of-fit test

- `library(lmtest)`

```
> AllData.test3 = lmer(MADRS.Raw.Score ~ Site + Condition + MADRS.Day * Median.RTCR + (1|Record.ID),
data=TwoBack, REML=FALSE)
```

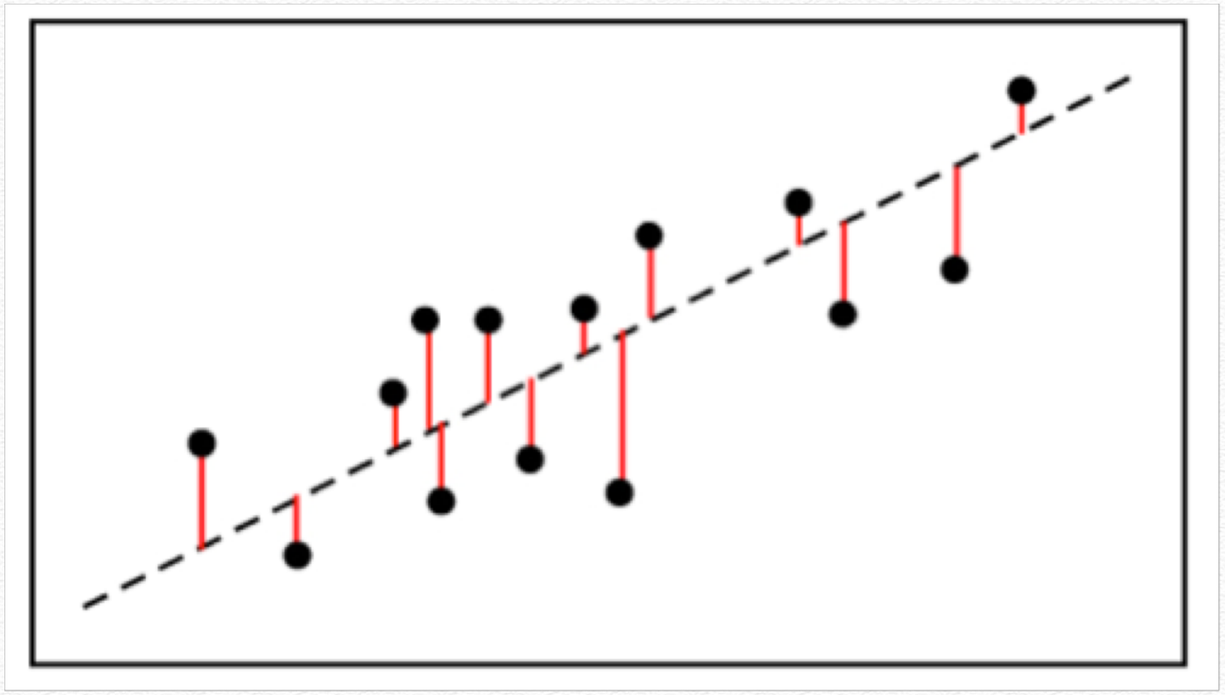
```
> lrtest(AllData.test2, AllData.test3)
Likelihood ratio test

Model 1: MADRS.Raw.Score ~ Site + Condition + MADRS.Day + Median.RTCR +
    (1 | Record.ID)
Model 2: MADRS.Raw.Score ~ Site + Condition + MADRS.Day * Median.RTCR +
    (1 | Record.ID)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   7 -392.20
2   8 -392.18  1 0.0284     0.8662
```

Coefficient of determination – R^2

- Proportion of the variance in the dependent variable that is predictable from the independent variable(s)

$$SS_{\text{res}} + SS_{\text{reg}} = SS_{\text{tot}}.$$

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$



Coefficient of determination – R^2

- library(MuMIn)

```
> r.squaredGLMM(AllData.test2)
          R2m       R2c
[1,] 0.2955092 0.7554878
```

- Marginal R^2 and describes the proportion of variance explained by the fixed factor(s) alone.
- Conditional R^2: proportion of variance explained by both the fixed and random factors.
- The expected value of R^2 depends on fields of study
- The higher, the better?
  - Don't overfit!

Missing Data

- Complete case analysis
  - May lost a lot of subjects
  - Assume completed cases are a random sample of the originally targeted sample.
- Multiple Imputation
  - Creates several different plausible imputed data sets
  - Each variable has its own imputation model
  - Appropriately combining results obtained from each imputed data set

Multiple Imputation

Predictive Mean Matching

- library(mice)

```
> madrs.imp = mice(MADRSData, m = 20, meth='pmm', print = FALSE)
```

- Convert mids object to a long-format data frame

```
> long_madrs = mice::complete(madrs.imp, action = 'long', include = TRUE)
```

	.imp	.id	Record.ID	Site	Condition	Baseline_Visit	CheckIn_2	CheckIn_1
17	0	17	P713	fMRI Guided	Task	33	27	28
18	0	18	P754	Standard (6cm)	Video	29	20	20
19	0	19	P805	fMRI Guided	Video	NA	NA	8
20	0	20	P842	Standard (6cm)	Task	37	7	0
21	0	21	P843	fMRI Guided	Task	33	26	20
22	0	22	P884	fMRI Guided	Video	26	3	NA
23	0	23	P908	fMRI Guided	Video	19	4	NA
24	0	24	P910	Standard (6cm)	Video	32	31	31
25	0	25	P941	fMRI Guided	Video	25	12	12
26	1	1	P109	fMRI Guided	Task	21	8	9
27	1	2	P115	fMRI Guided	Video	36	0	3
28	1	3	P133	Standard (6cm)	Video	25	13	16
29	1	4	P205	Standard (6cm)	Video	27	11	7
30	1	5	P225	Standard (6cm)	Task	37	20	28
31	1	6	P286	Standard (6cm)	Video	19	15	19
32	1	7	P323	Standard (6cm)	Task	24	2	1
33	1	8	P342	fMRI Guided	Video	31	21	20
34	1	9	P349	fMRI Guided	Task	37	29	19
35	1	10	P366	Standard (6cm)	Task	27	19	14
36	1	11	P404	Standard (6cm)	Task	29	29	19
37	1	12	P438	Standard (6cm)	Video	36	8	3
38	1	13	P521	Standard (6cm)	Video	24	20	17
39	1	14	P597	fMRI Guided	Task	33	35	22
40	1	15	P610	Standard (6cm)	Task	29	14	12
41	1	16	P626	fMRI Guided	Video	36	26	11
42	1	17	P713	fMRI Guided	Task	33	27	28
43	1	18	P754	Standard (6cm)	Video	29	20	20
44	1	19	P805	fMRI Guided	Video	25	11	8
45	1	20	P842	Standard (6cm)	Task	37	7	0
46	1	21	P843	fMRI Guided	Task	33	26	20
47	1	22	P884	fMRI Guided	Video	26	3	0
48	1	23	P908	fMRI Guided	Video	19	4	9
49	1	24	P910	Standard (6cm)	Video	32	31	31
50	1	25	P941	fMRI Guided	Video	25	12	12

Multiple Imputation

- Convert the manipulated data back to mids object

```
> madrs.imp2 = as.mids(long_madrs)
```

- Fit linear model to every imputed dataset

```
> mice_fit_madrs = with(madrs.imp2, lm(CheckIn_average ~ Site + Condition + Baseline_Visit +
SFNB2.SFNB_RTC2, data = MADRSData))
```

  - List object mira

- Combine the estimates from multiple imputation

```
> summary(pool(mice_fit_madrs))
```

  - mipo

Thank you!